What can computational models tell us about face processing?

Garrison W. Cottrell

Gary's Unbelievable Research Unit (GURU)

Computer Science and Engineering Department

Institute for Neural Computation

UCSD

Collaborators:

Matthew N. Dailey, Personal Robotics, Inc.

Maki Sugimoto, HNC, Inc.

Ralph Adolphs, University of Iowa Department of Neurology

Curtis Padgett, NASA JPL

And now…Lingyun Zhang, Matt Tong, Zak Haque, Honghao Shan, Jonathan Nelson, Nam Nguyen, Brian Tran, Brenden Lake

How (I like) to build Cognitive Models

- I like to be able to relate them to the brain, so “neurally plausible” models are preferred -- neural nets.
- The model should be a *working model* of the *actual* task, rather than a cartoon version of it.
- Of course, the model should nevertheless be *simplifying* (i.e. it should be constrained to the essential features of the problem at hand).
- Then, take the model “as is” and fit the experimental data: 0 fitting parameters is preferred over 1, 2 , or 3.

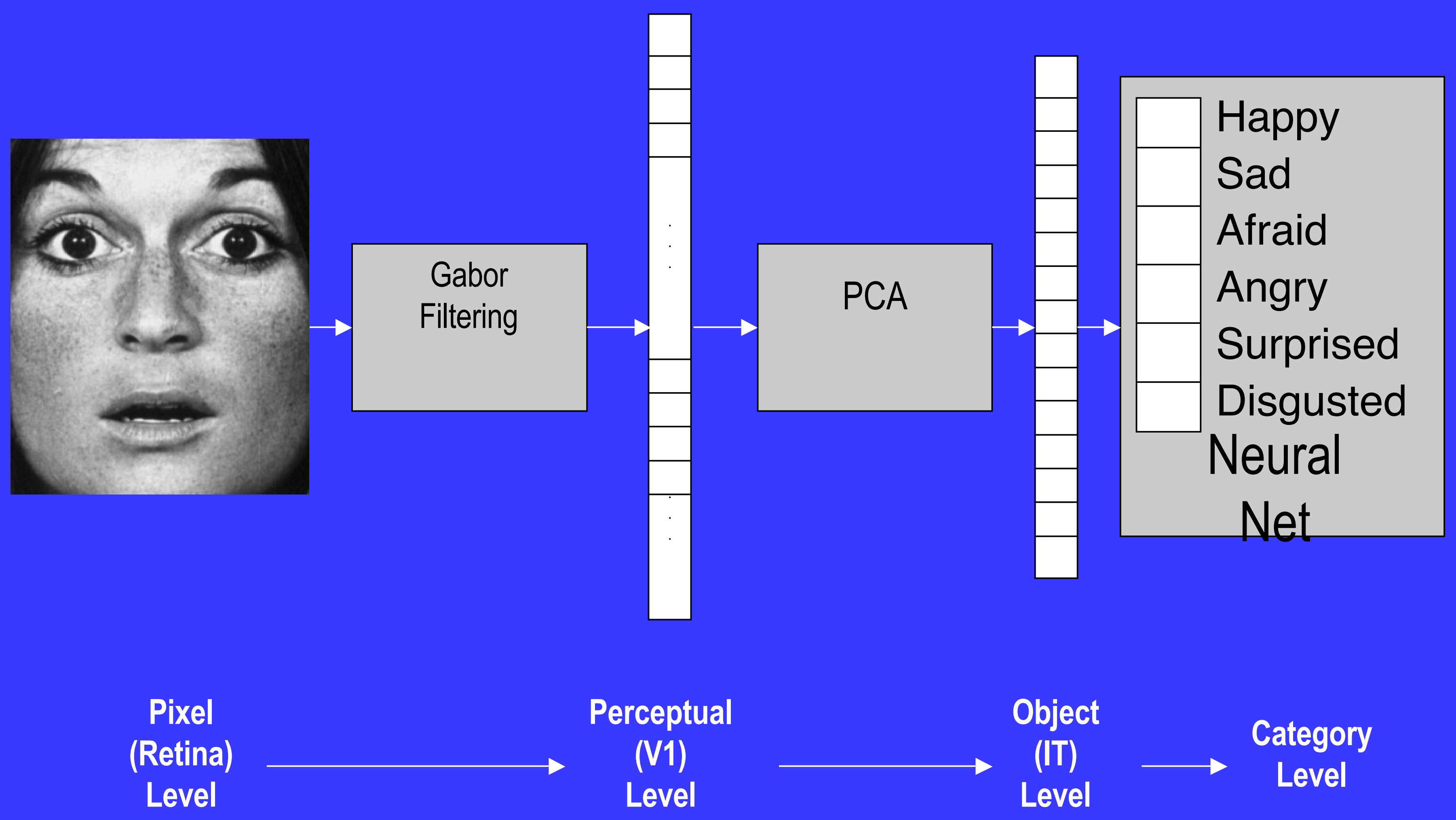
The *other* way (I like) to build Cognitive Models

- Same as above, except:
- Use them as *exploratory* models -- in domains where there is little direct data (e.g. no single cell recordings in infants or undergraduates) to suggest what we might find if we *could* get the data. These can then serve as “intuition pumps.”
- Examples:
  - Why we might get specialized face processors
  - Why those face processors get recruited for other tasks

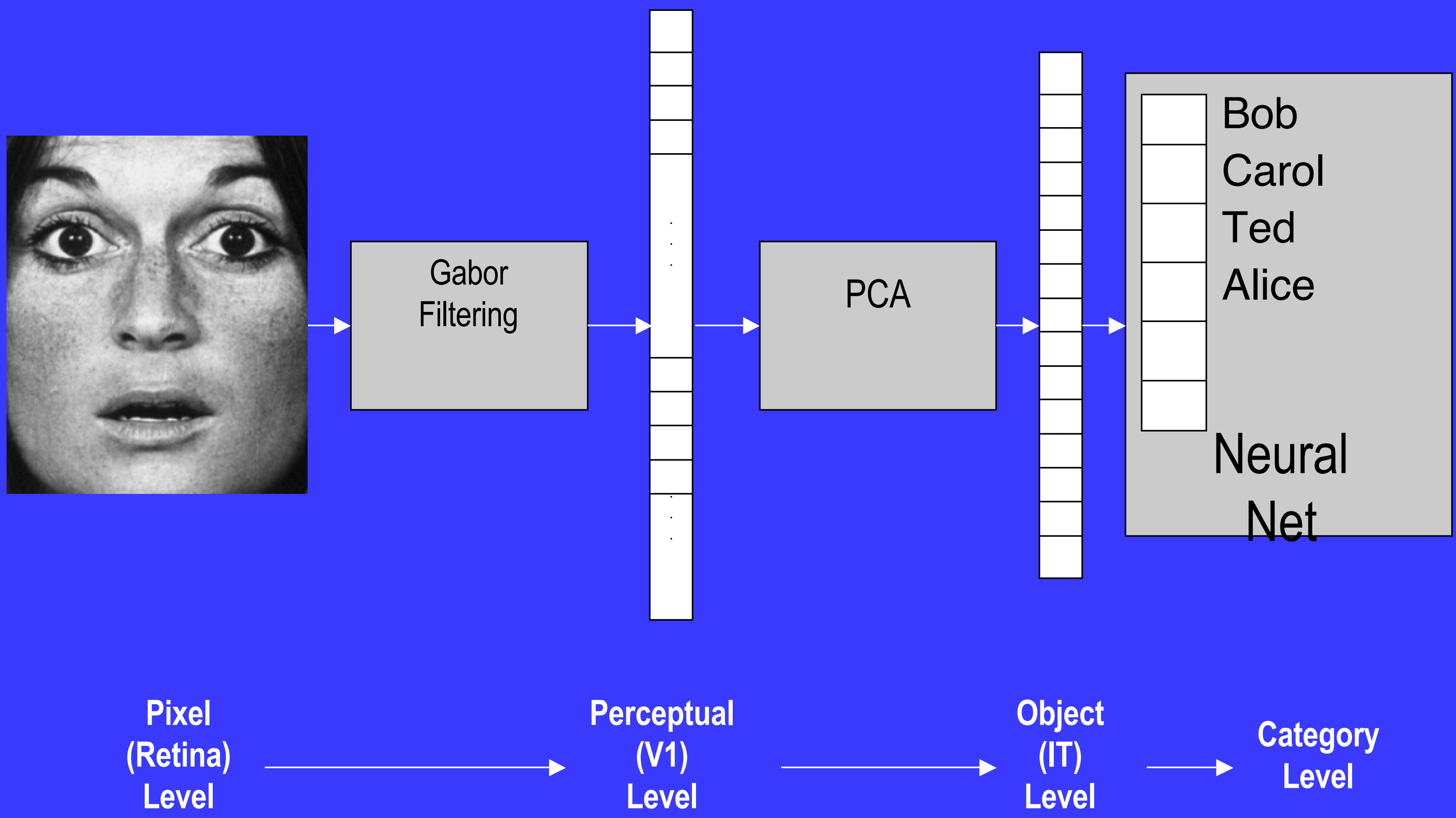
Outline

- Review of our model(s) of face (and object) classification.
- (Very brief!) summary of results in face specialization (exploratory model)
- Summary of results in expression recognition (data fitting model)
- Tour of our model of visual expertise (exploratory model)
- Wrap up

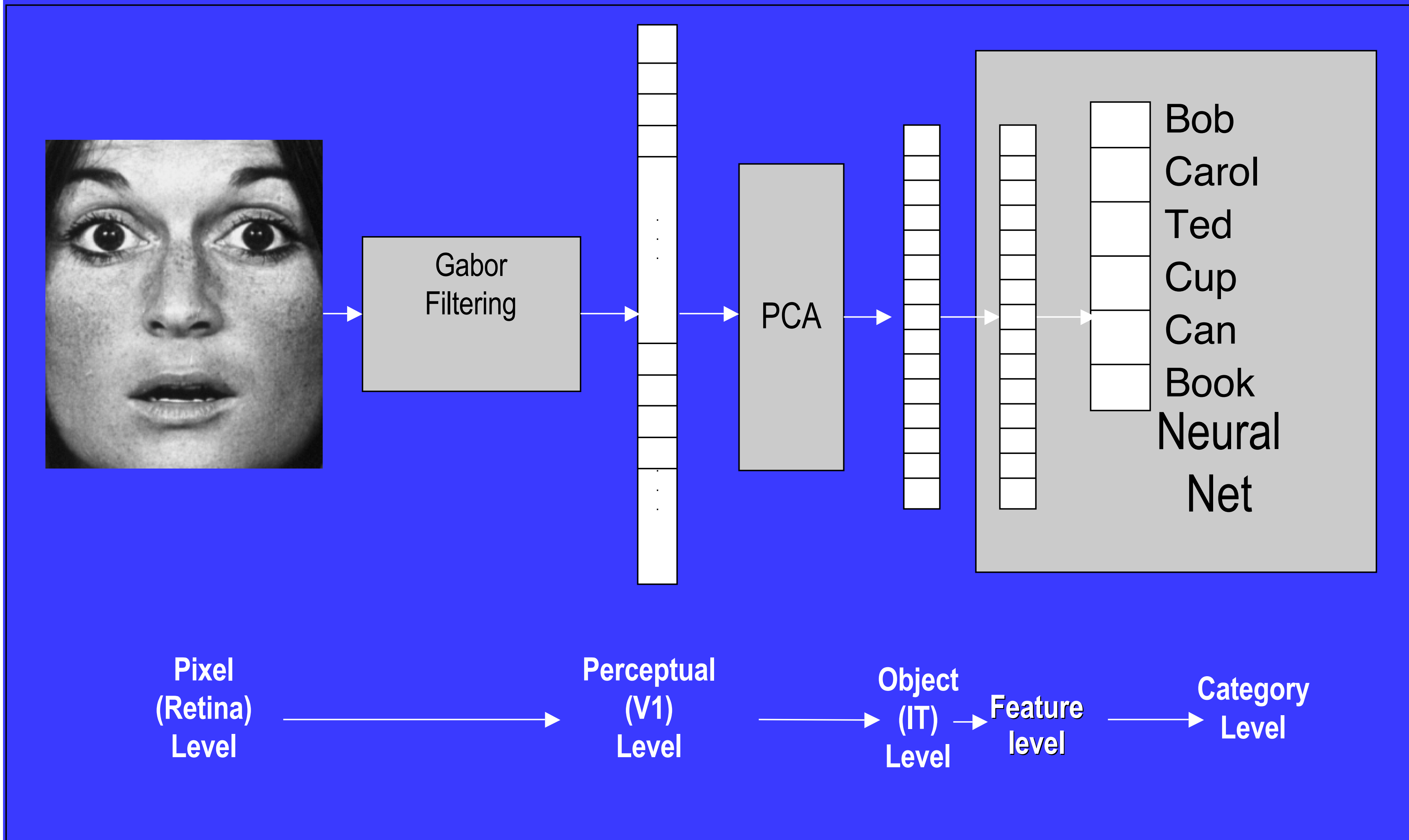
The Face Processing System



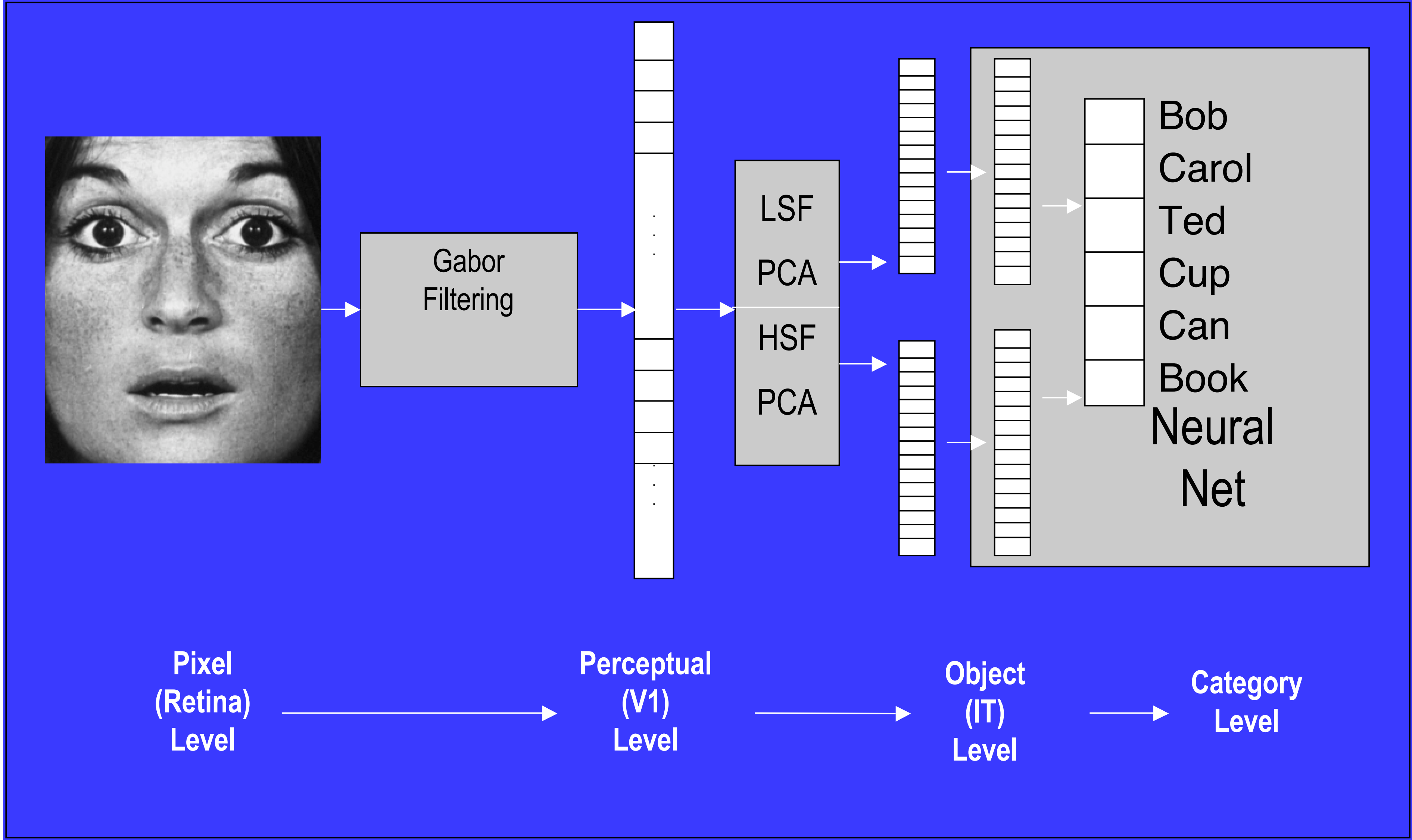
The Face Processing System



The Face Processing System

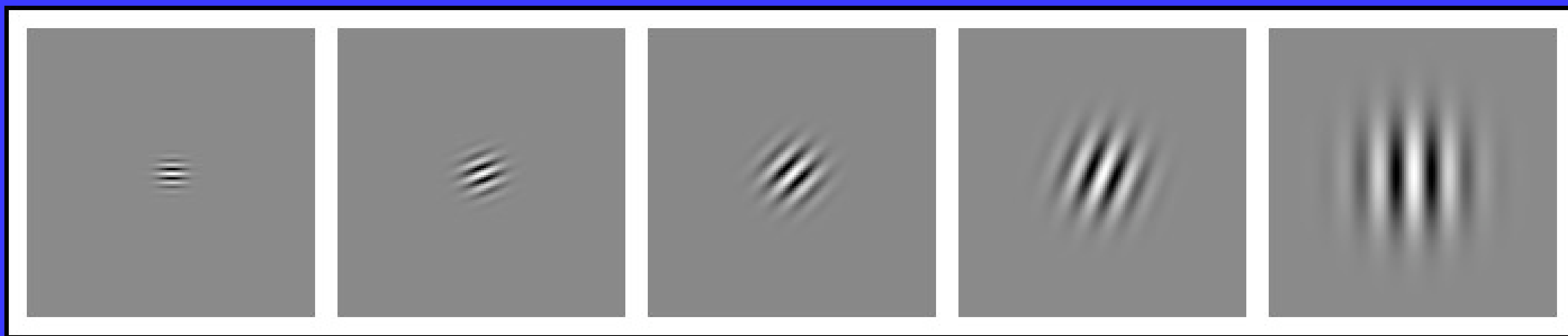


The Face Processing System

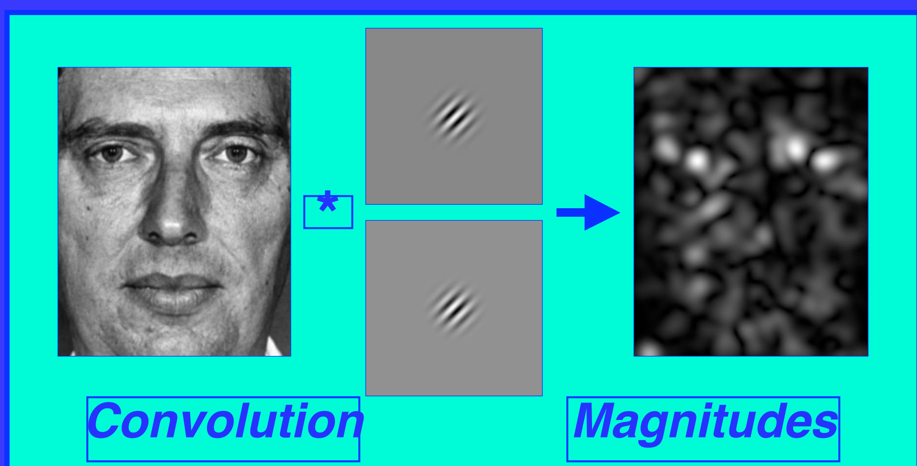


The Gabor Filter Layer

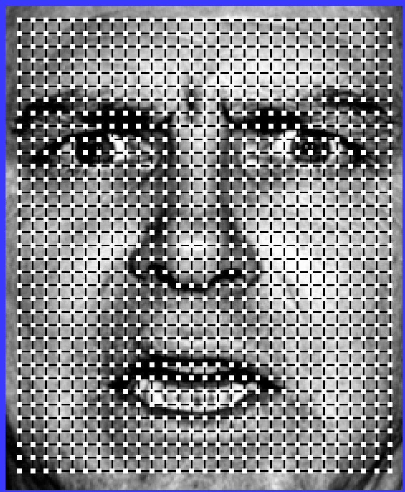
- Basic feature: the 2-D Gabor wavelet filter (Daugman, 85):



- These model the processing in early visual areas

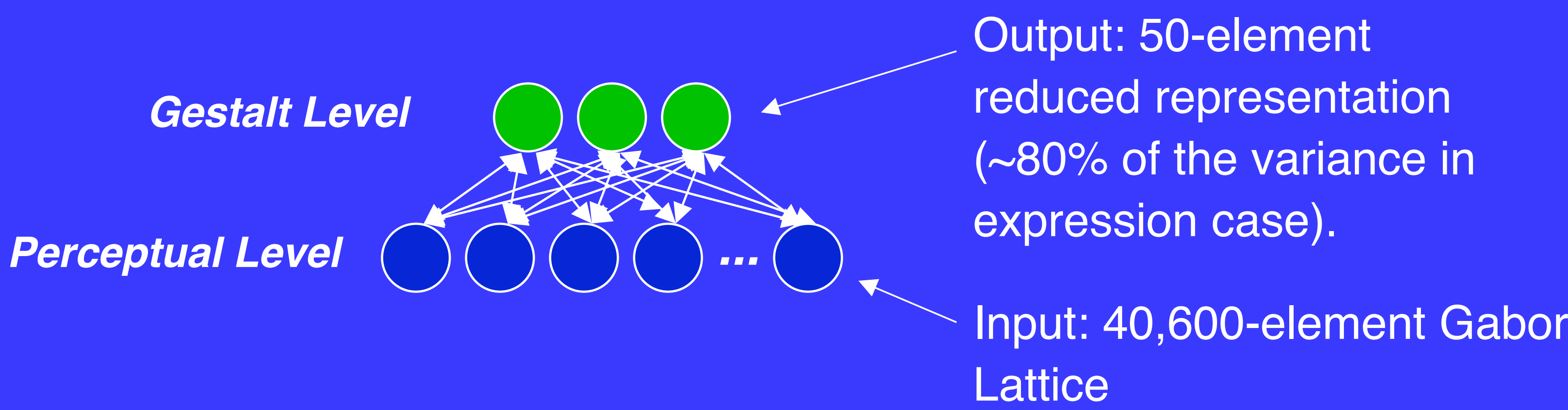


Subsample in a 29x36 grid



The Gestalt Level

- We reduce dimensionality of the perceptual-level representation with Principal Components Analysis (PCA):



- This is neurologically plausible because PCA can be learned by Hebbian networks.
- The resulting 50-element vector is input to the category level.

The Final Layer: Classification

- The final layer is trained based on the category of the stimulus: expression, identity, object class - one output per class.
- Categories can be at different levels: basic, subordinate.
- Simple learning rule (~delta rule). It says (mild lie here):
  - ***add*** inputs to your weights (synaptic strengths) when you are supposed to be ***on***,
  - ***subtract*** them when you are supposed to be ***off***.
- This makes your weights "look like" your favorite patterns – the ones that turn you on.
- When no hidden units => No back propagation of error.
- When hidden units: we get task-specific features (most interesting when we use the basic/subordinate distinction)

Correlates to Psychological Variables

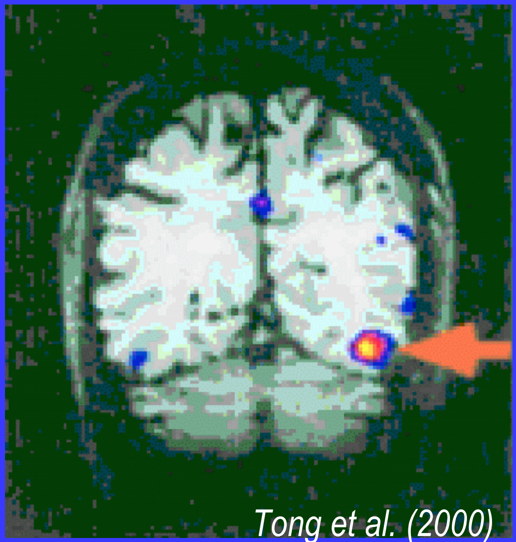
- 1 trained neural network = 1 human subject.
- "Answer" (button push, etc.) = highest network output
- Response distribution = average over multiple network outputs
- Response time = uncertainty of maximal output (1.0 - y_{max}).
- Errors: Errors! I.e., when highest output is wrong answer
- Similarity: correlation between representations at a particular level of processing (note: best fitting level => suggestion that we use that level)
- Discriminability: 1 – similarity

Outline

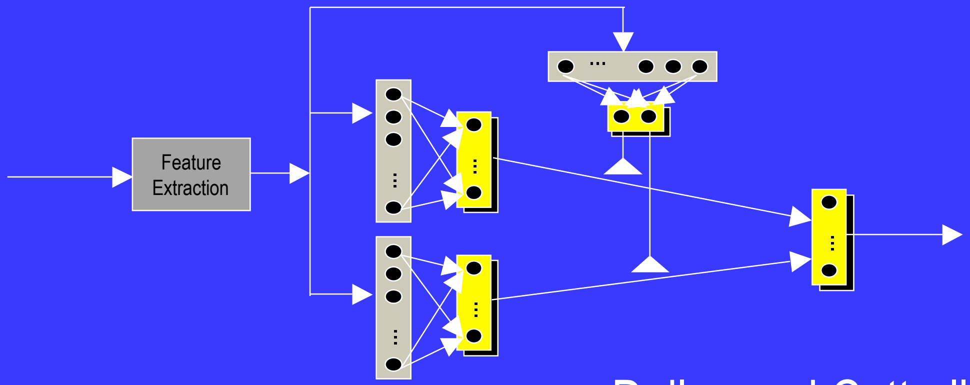
- Review of our model(s) of face (and object) classification.
- (Very brief!) summary of results in face specialization (exploratory model)
- Summary of results in expression recognition (data fitting model)
- Tour of our model of visual expertise (exploratory model)
- Wrap up

Face Specialization

- Why do we have a face processor in fusiform gyrus? Our model suggests that there is an interaction between
  - **Low spatial frequency (LSF) information and**
  - **The task of face expertise (subordinate level categorization)**
- Given competing networks, the one that gets the LSF's wins
- Recent behavioral, fMRI and ERP data support this account (Schyns & Oliva, 1999; Gauthier et al. 1999; Goffaux et al., 2002)



Tong et al. (2000)



Dailey and Cottrell (1999), *Neural Networks*.

Outline

- Review of our model(s) of face (and object) classification.
- (Very brief!) summary of results in face specialization (exploratory model)
- Summary of results in expression recognition (data fitting model)
- Tour of our model of visual expertise (exploratory model)
- Wrap up

The Issue: Are Similarity and Categorization Two Sides of the Same Coin?

- Some researchers believe perception of facial expressions is a new example of ***categorical perception:***
  - Like the colors of a rainbow, the brain separates expressions into discrete categories, with:
  - Sharp boundaries between expressions, and…
  - Higher discrimination of faces near those boundaries.

The Issue: Are Similarity and Categorization Two Sides of the Same Coin?

- Some researchers believe the underlying representation of facial expressions is NOT discrete:
  - There are two (or three) underlying dimensions, e.g., intensity and valence (found by MDS).
  - Our perception of expressive faces induces a *similarity structure* that results in a circle in this space
- Our model of expression recognition accounts for both kinds of data

Expression recognition

- Here, we trained a simple neural network to classify the six "basic" facial expressions, using the Ekman & Friesen "Pictures of Facial Affect" (POFA) database.
- We fit (without fit parameters) a variety of data consistent with:
  - The "discrete categories" account of facial expression recognition (the categorical perception account).
  - The "continuous, multidimensional space" account of facial expression perception (the "emotion circumplex" account).
- Hence, these data need not be at odds (but the discrete folks need to rethink their position).

Dailey, Cottrell, Padgett, and Adolphs (2002), *Journal of Cognitive Neuroscience*

Facial Expression Database

- Ekman and Friesen quantified muscle movements (Facial Actions) involved in prototypical portrayals of happiness, sadness, fear, anger, surprise, and disgust.
  - Result: the Pictures of Facial Affect Database (1976).
  - 70% agreement on emotional content by naive human subjects.
- 110 images, 14 subjects, 7 expressions.

Anger, Disgust, Neutral, Surprise, Happiness (twice), Fear, and Sadness
This is actor "JJ": The easiest for humans (and our model) to classify

Results (Generalization)

Expression	*Network % Correct*	*Human % Agreement*
Happiness	*100.0%*	*98.7%*
Surprise	*100.0%*	*92.4%*
Disgust	*100.0%*	*92.3%*
Anger	*89.2%*	*88.9%*
Sadness	*82.9%*	*89.2%*
Fear	*66.7%*	*87.7%*
Average	*89.9%*	*91.6%*

- Kendall's τ (rank order correl.) of emotion difficulty: .667, p=.0441
- Fear is hard because it is the most confusable expression.

Examining the Net’s Representations

- We want to visualize “receptive fields” in the network.
- But the Gabor magnitude representation is noninvertible.
- We can *learn* an approximate inverse mapping, however.
- We used linear regression to find the best linear combination of Gabor magnitude principal components for each image pixel.
- Then projecting each unit’s *weight vector* into image space with the same mapping visualizes its “receptive field.”

Examining the Net's Representations

- The "y-intercept" coefficient for each pixel is simply the average pixel value at that location over all faces, so subtracting the resulting "average face" shows more precisely what the units attend to:

- Apparently local features appear in the global templates.

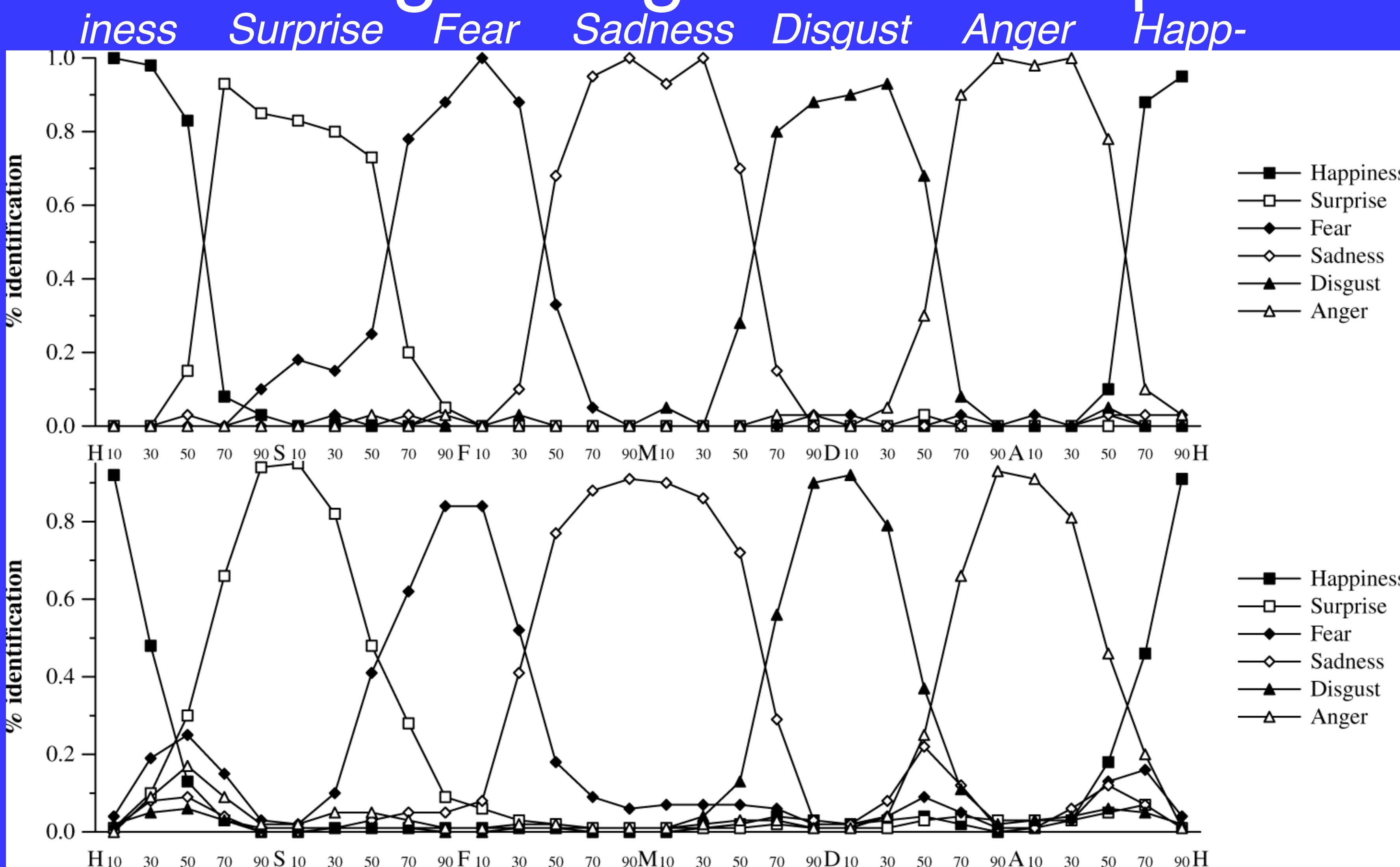
Morph Transition Perception

- Morphs help psychologists study categorization behavior in humans
- Example: JJ Fear to Sadness morph:

0% 10% 30% 50% 70% 90% 100%

- Young et al. (1997) Megamix: presented images from morphs of all 6 emotions (15 sequences) to subjects in random order, task is 6-way forced choice button push.

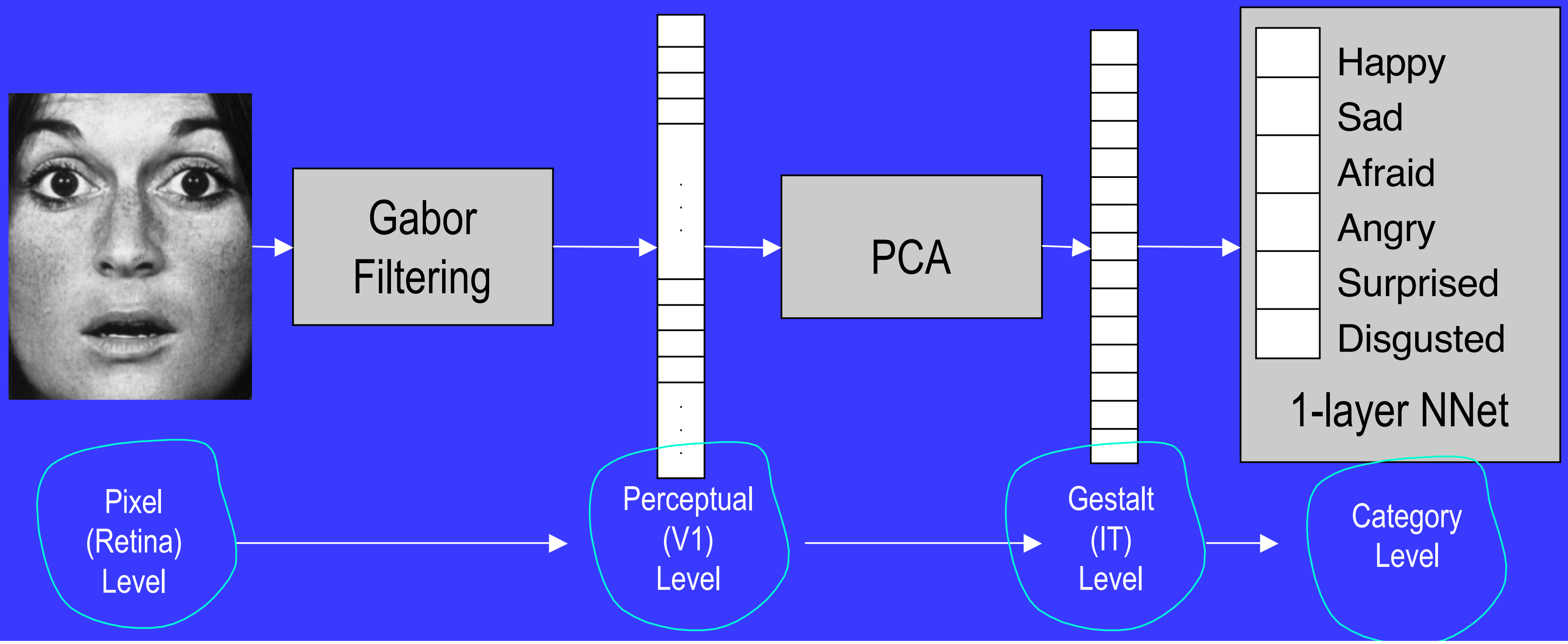
Modeling Categorical Perception



- Overall correlation r=.9416, with NO FIT PARAMETERS!

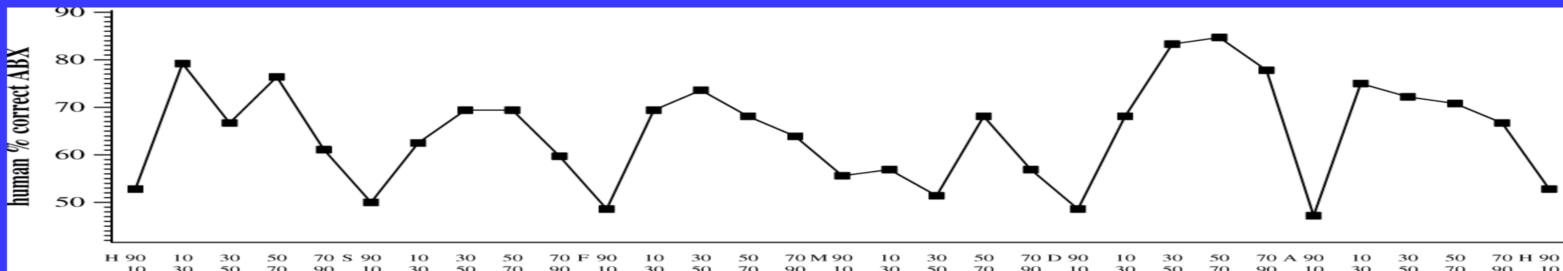
Modeling Discrimination (for CP)

- Is improved discrimination near boundaries due to influence of the categories?
- Discrimination is naturally modeled as the flip side of similarity:
  - We model discrimination as 1-*r* (correlation) between pairs.
- Prediction of CP: best fit should occur at category level of the model.

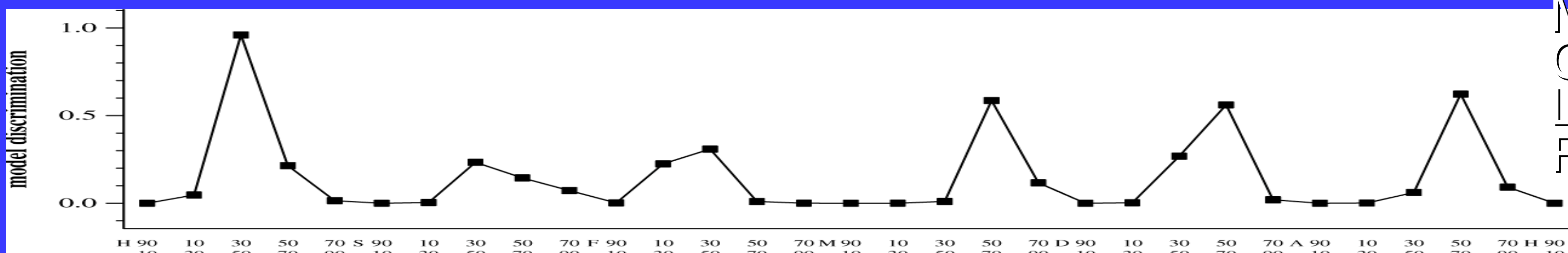


Model Discrimination Scores

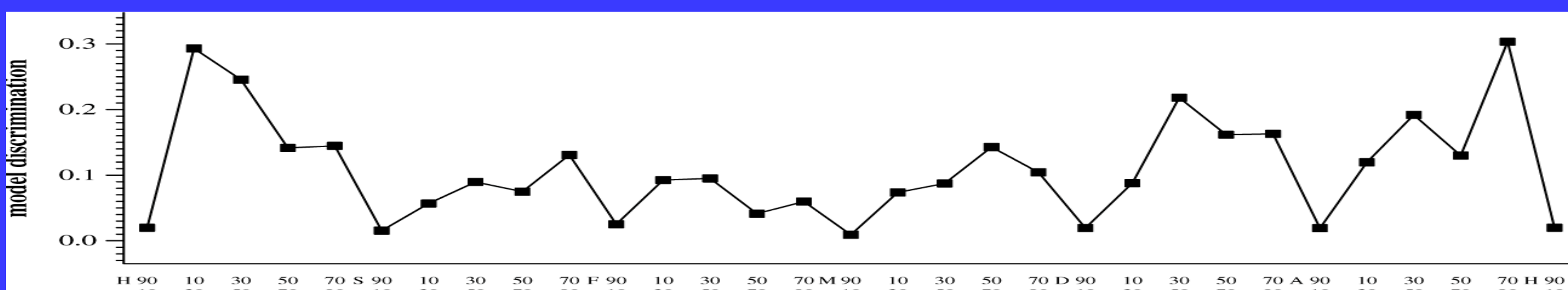
iness Surprise Fear Sadness Disgust Anger Happ-



HUMAN



MODEL OUTPUT LAYER R=0.36

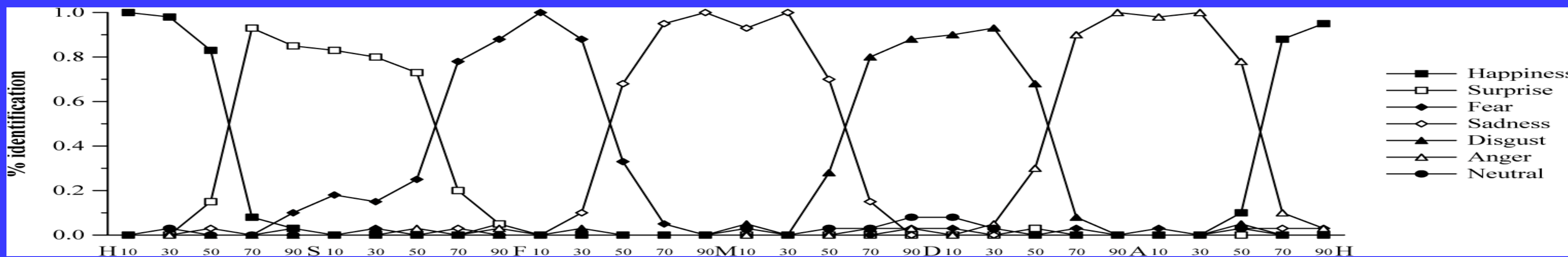


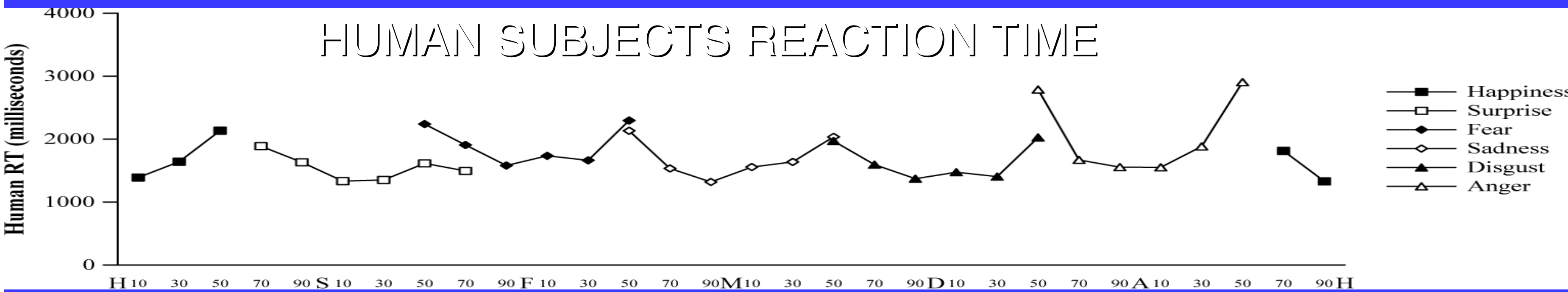
MODEL GESTA LAYER R=0.61

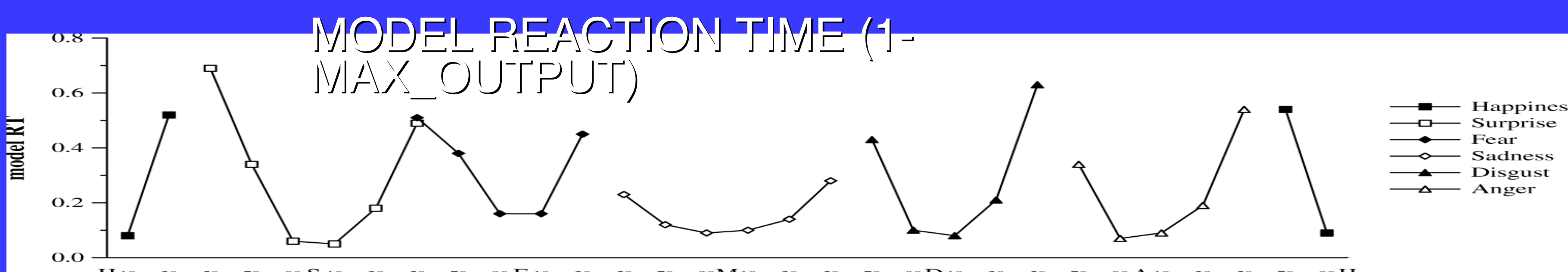
- The model fits the data best at a precategorical layer: The layer we call the “gestalt” layer; NOT at the category level

Non-CP effect 1: Reaction Time

iness   Surprise   Fear   Sadness   Disgust   Anger   Happ-







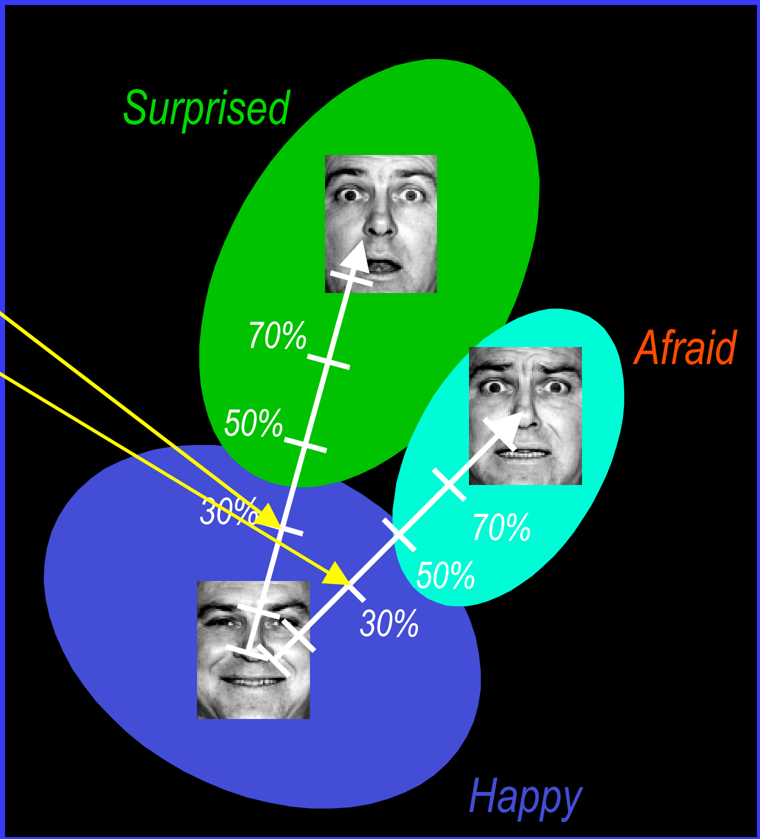
Correlation between model & data: .6771, $p<.001$

Non-CP effect 2: Detecting a Morph Trajectory

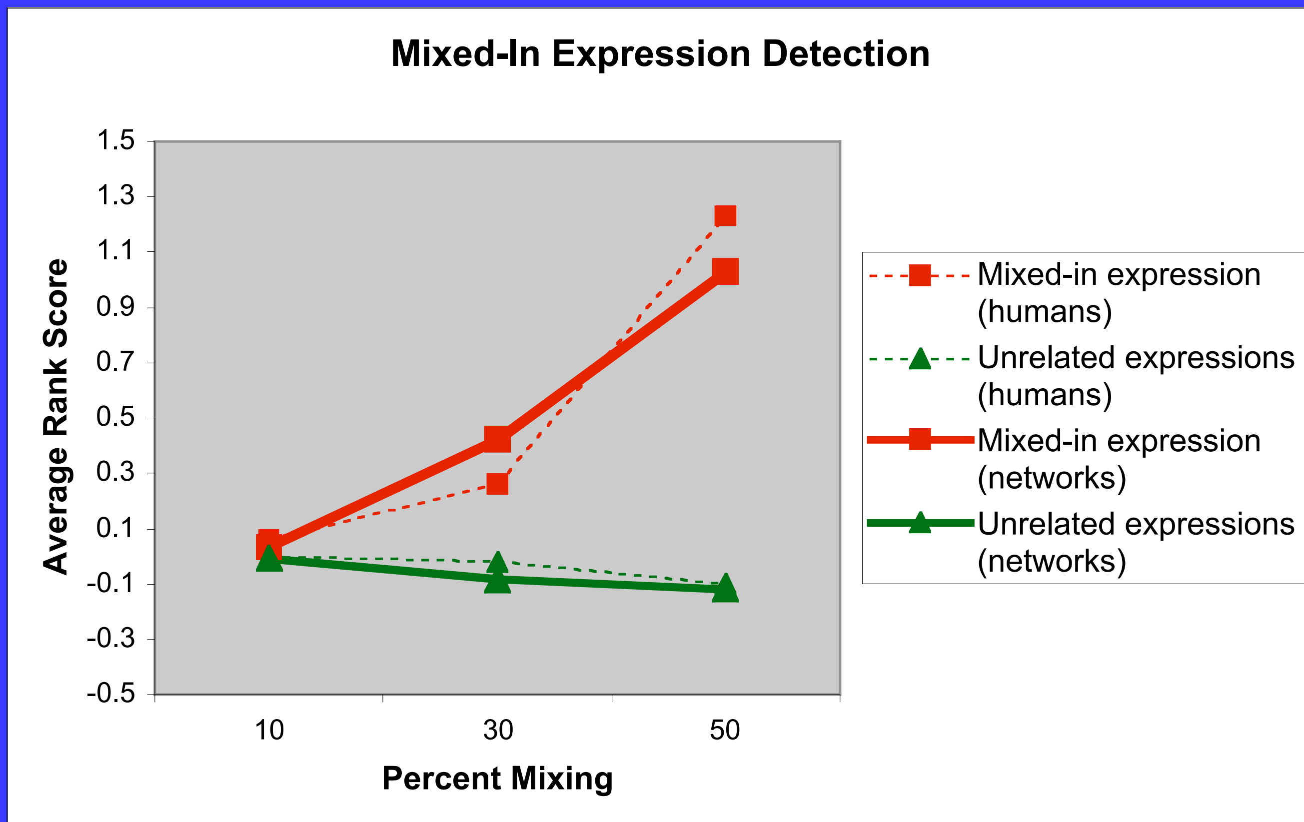
- A strong discrete categories theory would predict no perception of the structure internal to a category.

- But subjects are above chance at detecting the target emotion of 30% morphs!

- The model's sensitivity is **nearly identical** to human sensitivity within categories.



Mixed-in Expression Detection

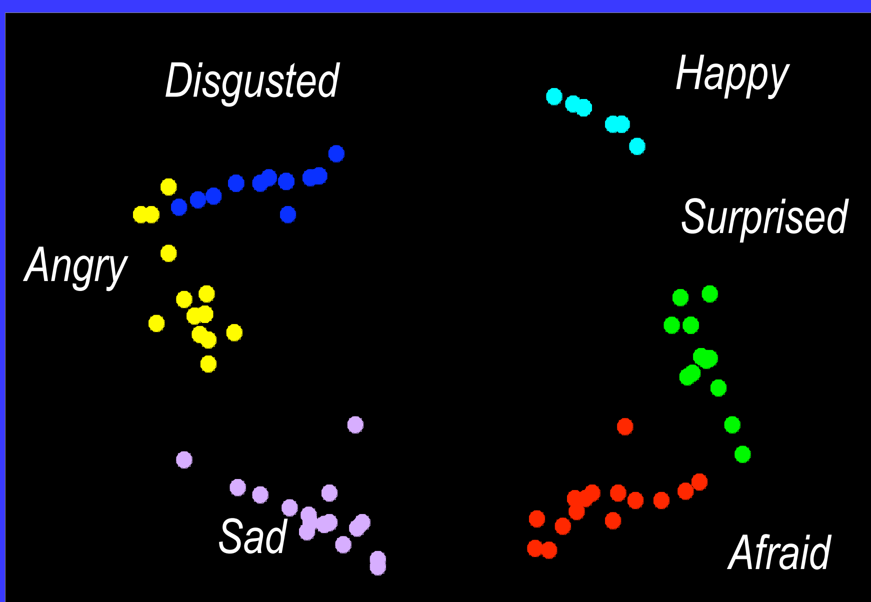


This analysis is based on the same measures used by Young et al. on the original human data.

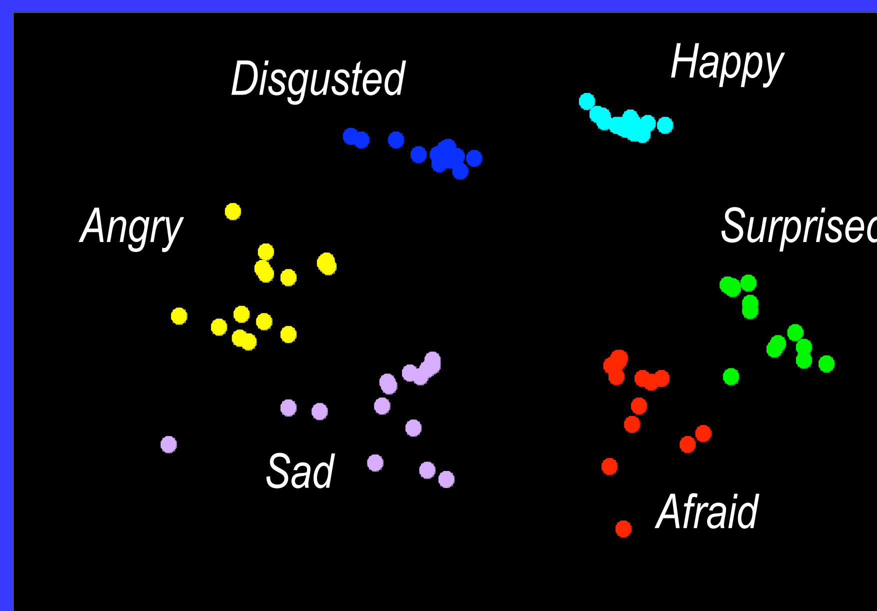
Similarity Structures

- Multidimensional scaling (MDS) helps visualize similarity ratings. The technique makes facial expression space look continuous.
- Human and model confusions lead to similar structures.
- Confusion matrices are also highly correlated on train and test sets.

Human



Model



Outline

- Review of our model(s) of face (and object) classification.
- (Very brief!) summary of results in face specialization (exploratory model)
- Summary of results in expression recognition (data fitting model)
- Tour of our model of visual expertise (exploratory model)
- Wrap up

Are you a perceptual expert?

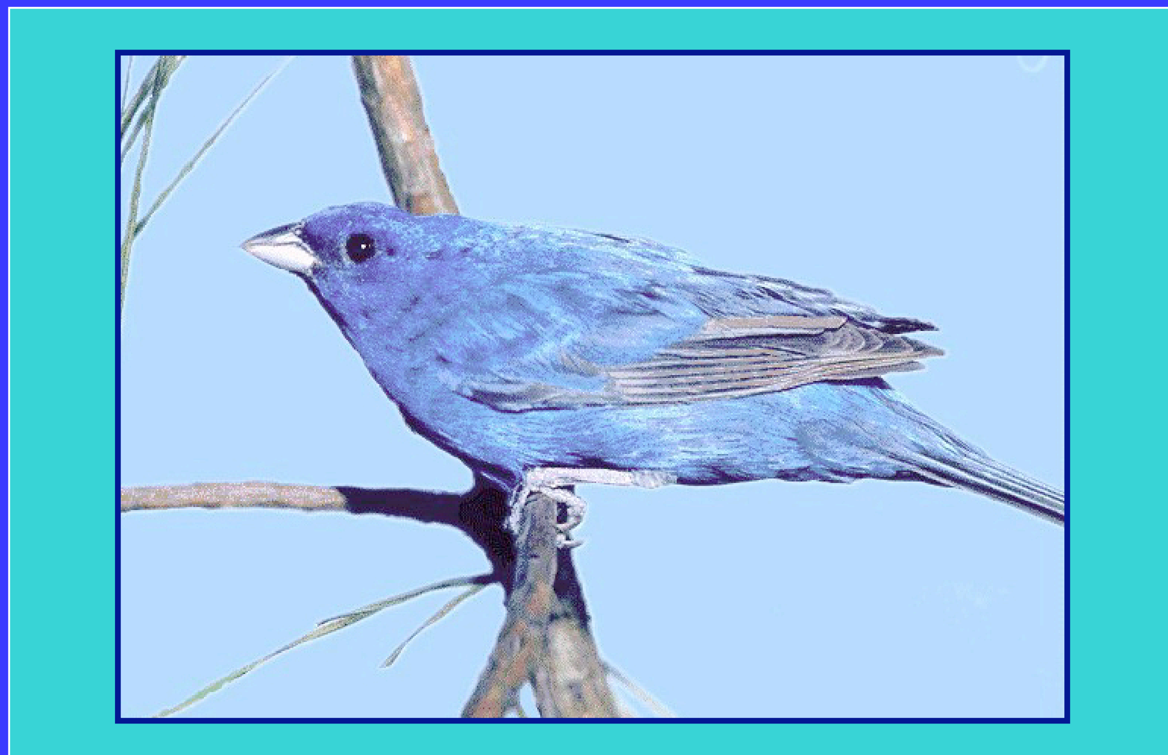
Take the expertise test!!!*

"Identify this object with the first name that comes to mind."

**Courtesy of Jim Tanaka, University of Victoria

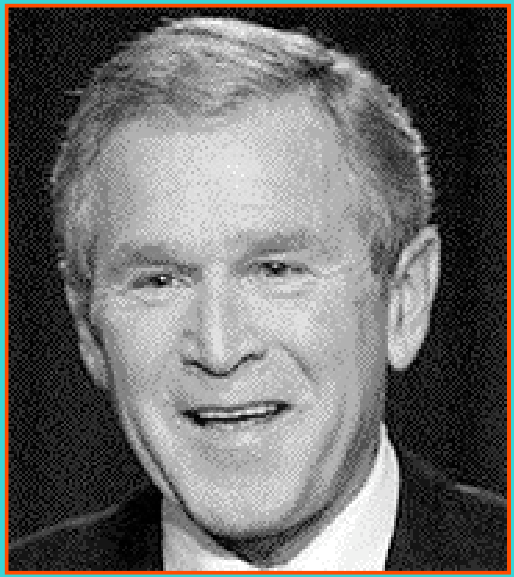
"**Car**" **-** Not an expert

"2002 BMW Series 7" - *Expert!*



“Bird” or “Blue Bird” - Not an expert

“Indigo Bunting” - ***Expert!***



“Face” or “Man” - Not an expert

“George Dubya”- *Expert!*

“Jerk” or “Megalomaniac” - Democrat!

How to identify an expert?

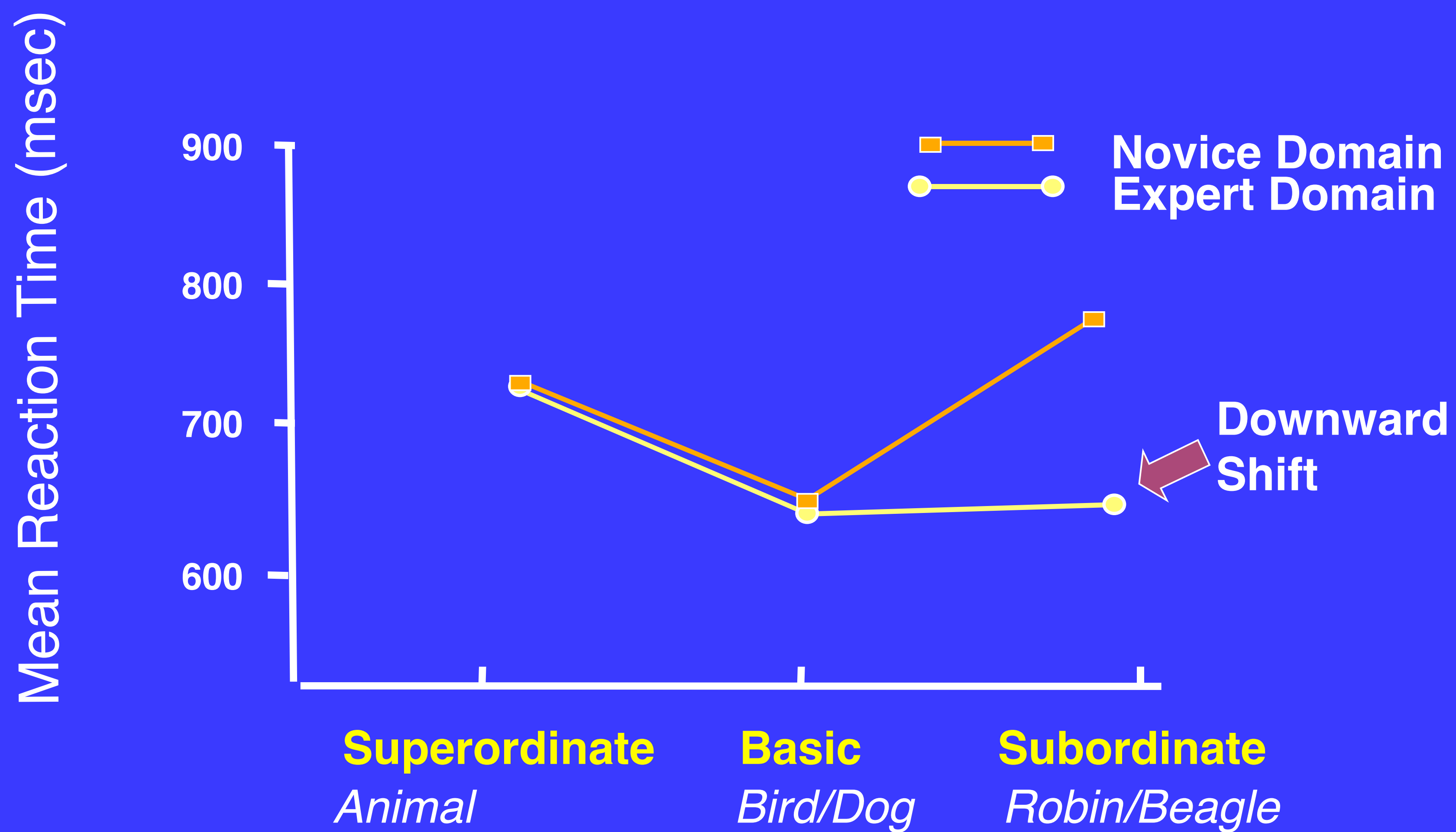
<u>Behavioral benchmarks of expertise</u>

- Entry level shift - can recognize items on category and individual level equally fast

<u>Neurological benchmarks of expertise</u>

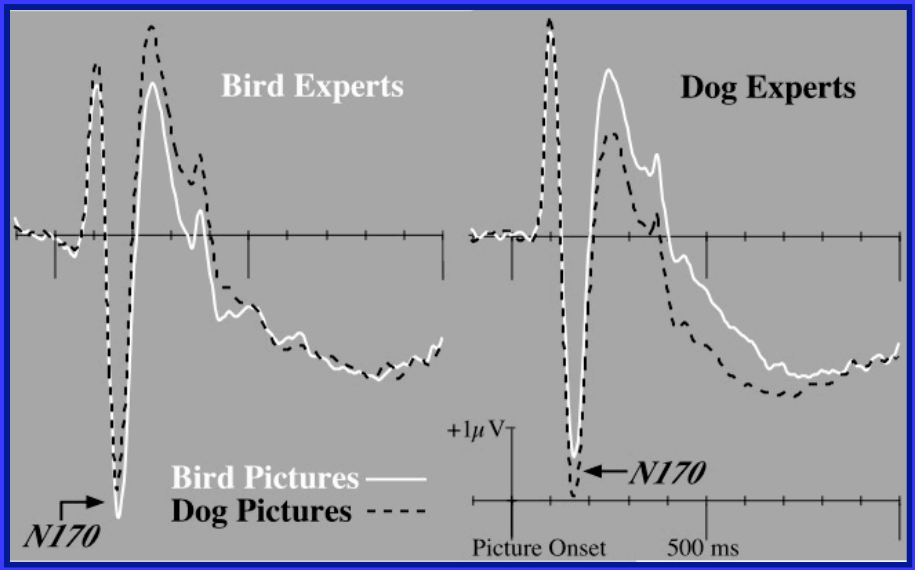
- Enhancement of N170 ERP brain component
- Increased activation of fusiform gyrus

Entry-Level Shift in Expertise

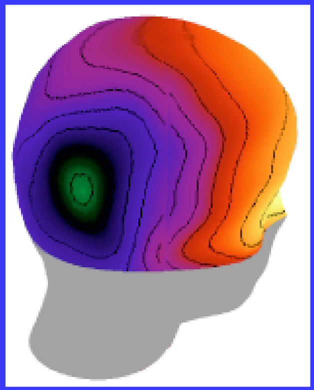


Neurologic Markers of Expertise

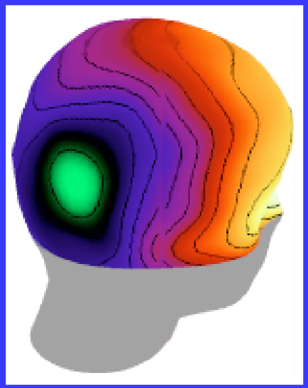
Event-related Potentials



Tanaka & Curran, 2001; see also Gauthier, Curran, Curby & Collins, 2003, *Nature Neuro*.



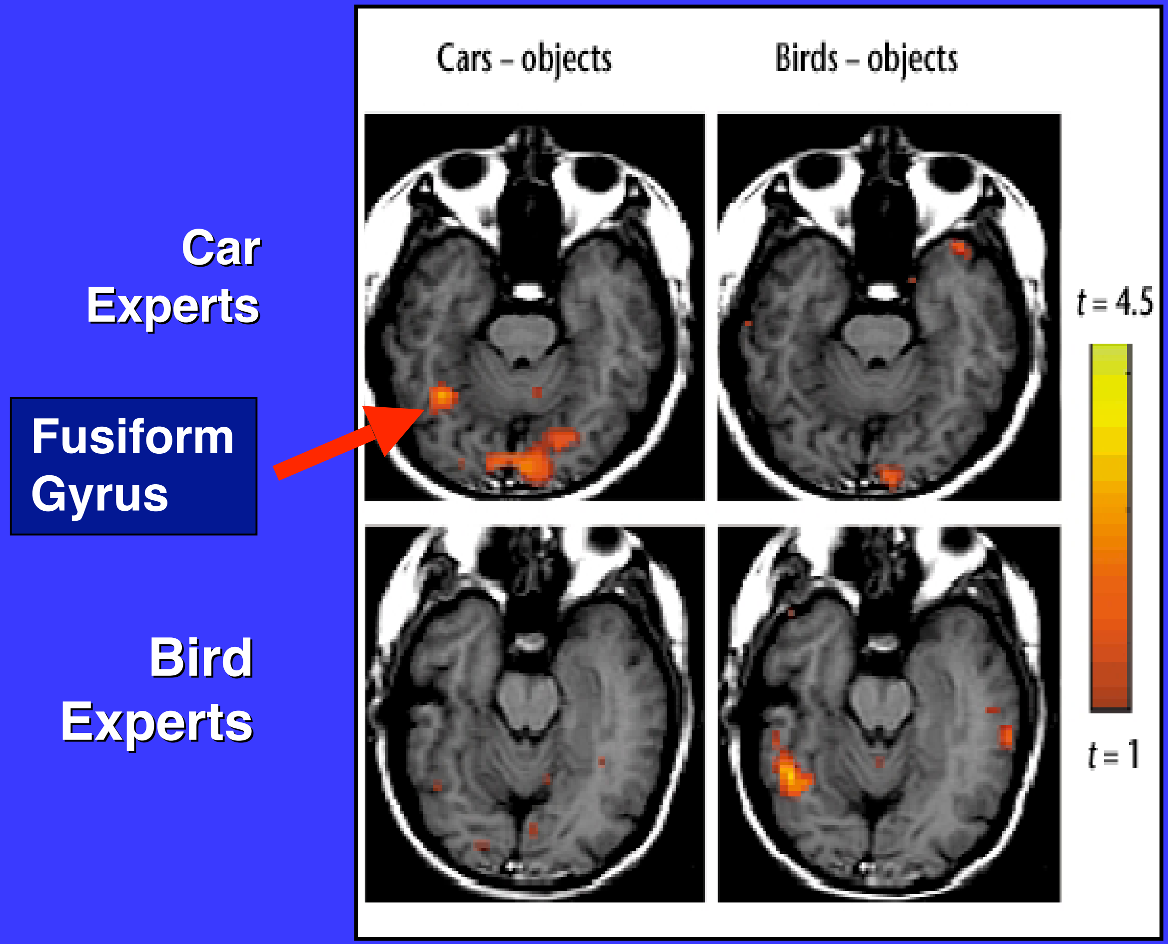
Novice Domain



Expert Domain

Neurologic Markers of Expertise

Neuroimaging



Visual expertise

- The so-called “Fusiform Face Area” (FFA) is apparently specialized for face processing.
- However, Gauthier and colleagues have shown that it *also* lights up for cars when the subject is a car expert, birds when the subject is a bird expert, Greebles when the subject is a Greeble expert (what’s a Greeble? Later.)
- Hence her view is that the FFA is an area associated with a *process*: fine level discrimination of homogeneous categories.
- But why would an area that presumably starts as a face area get recruited for these other visual tasks? Surely, they don’t share features, do they?

Sugimoto & Cottrell (2001), *Proceedings of the Cognitive Science Society*

Motivation: Evidence for the Face Specific View

- *Prosopagnosia* patients have a deficit in identifying individual faces but normal in detecting faces or other non-face objects, while *visual object agnosia* patients may be normal with face recognition but have a deficit in object recognition.
- fMRI shows the fusiform face area "lights up" for faces but not for objects (Kanwisher)
- Recognition of faces is more sensitive to configural changes than objects.

→ Face and non-face objects have separate processing mechanisms

Motivation: Evidence against the face specific view

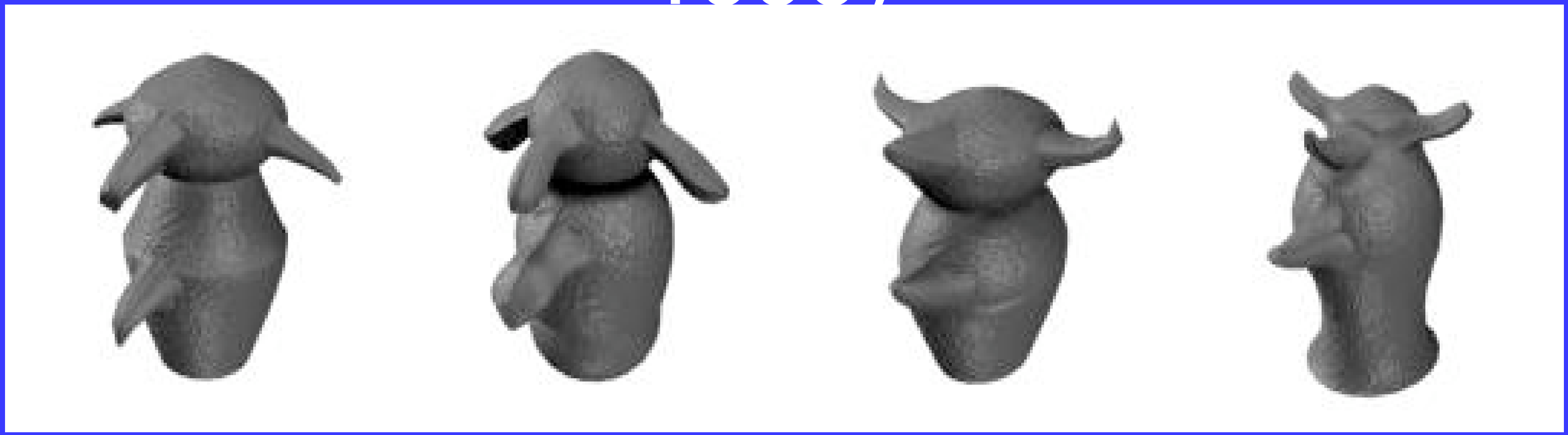
- Gauthier et al. point out that faces and objects differ not only in their image geometries, but also in …

1. Level of discrimination
2. Level of experience

➡ We are all face “experts”.

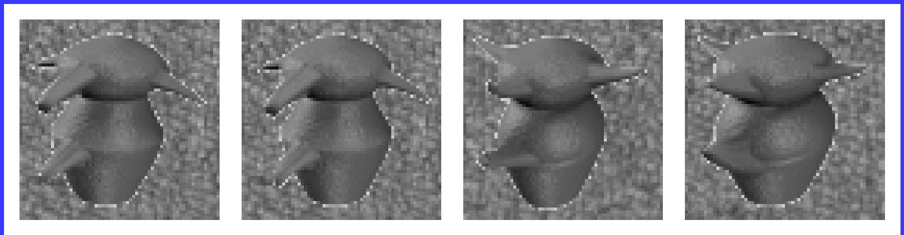
- FFA shows high activation for a wide variety of non-objects when these two conditions are controlled.

Greeble Experts (Gauthier et al. 1999)



- Subjects trained over many hours to recognize individual Greebles.
- Activation of the FFA increased for Greebles as the training proceeded.

Model Database



- 64x64 8bit grayscale images of faces, books, cups, cans and Greebles
- 12 individuals per category
- 5 different images per individual
- Total of 5x12x5=300 images

Main idea: We will pretrain at different levels of categorization.

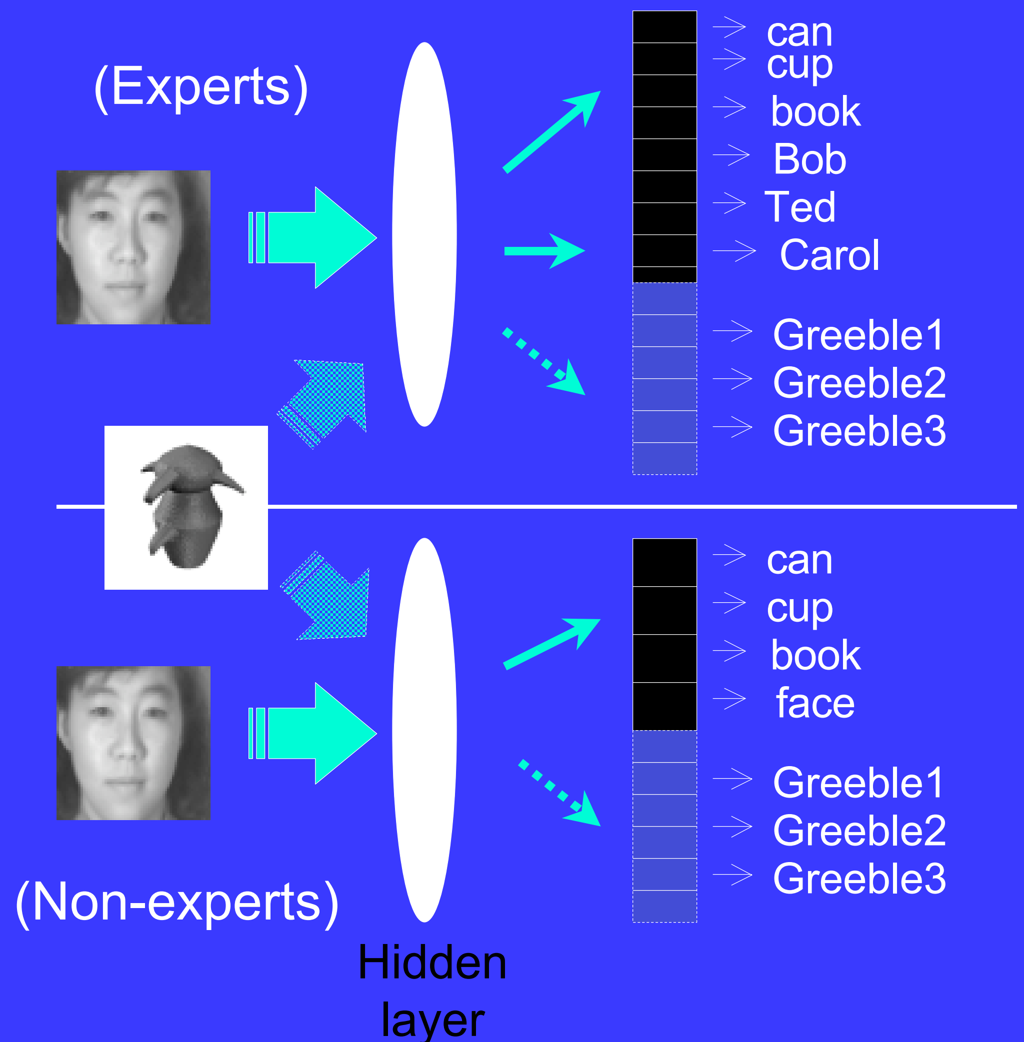
An "expert" is a network trained to individuate individuals.

A non-expert is a network trained only to categorize at the superordinate level.

Can an expert network learn the Greebles better?

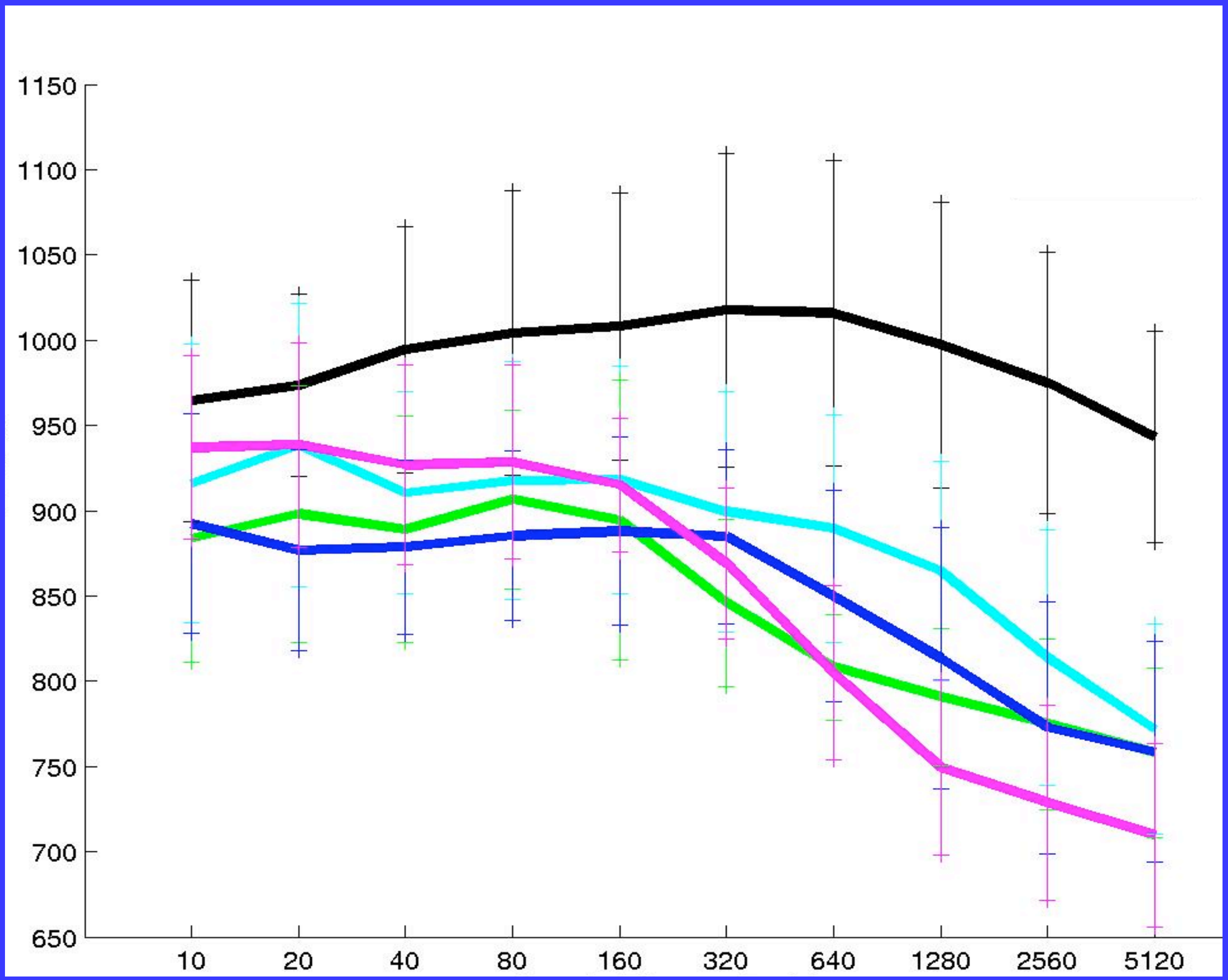
Model

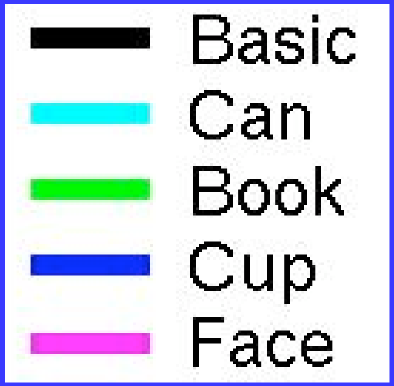
- Pretrain two groups of neural networks on different tasks.
- Compare the abilities to learn a new individual Greeble classification task.



Experts Learn Greebles Faster

Time to learn Greebles





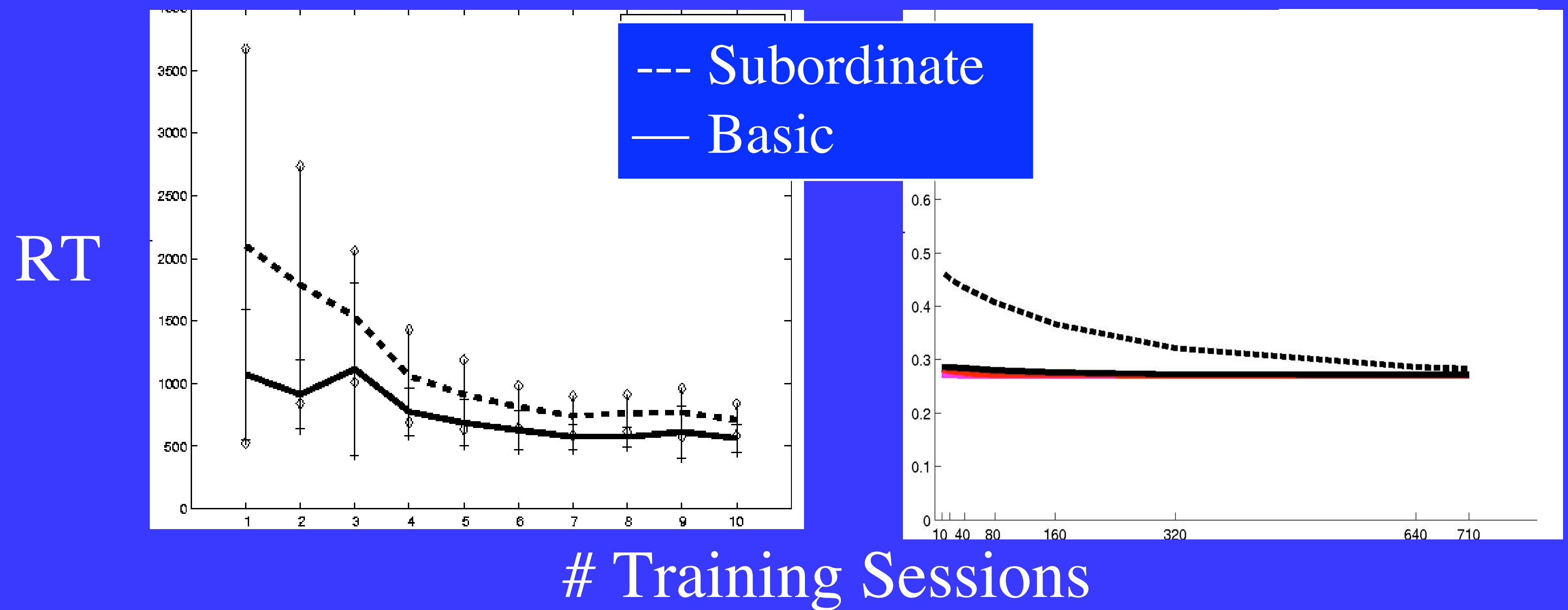
Training time on previous task

Entry Level Shift: Subordinate RT decreases with training

(rt = uncertainty of response = 1.0 -*max(output)*)

Human data

Network data



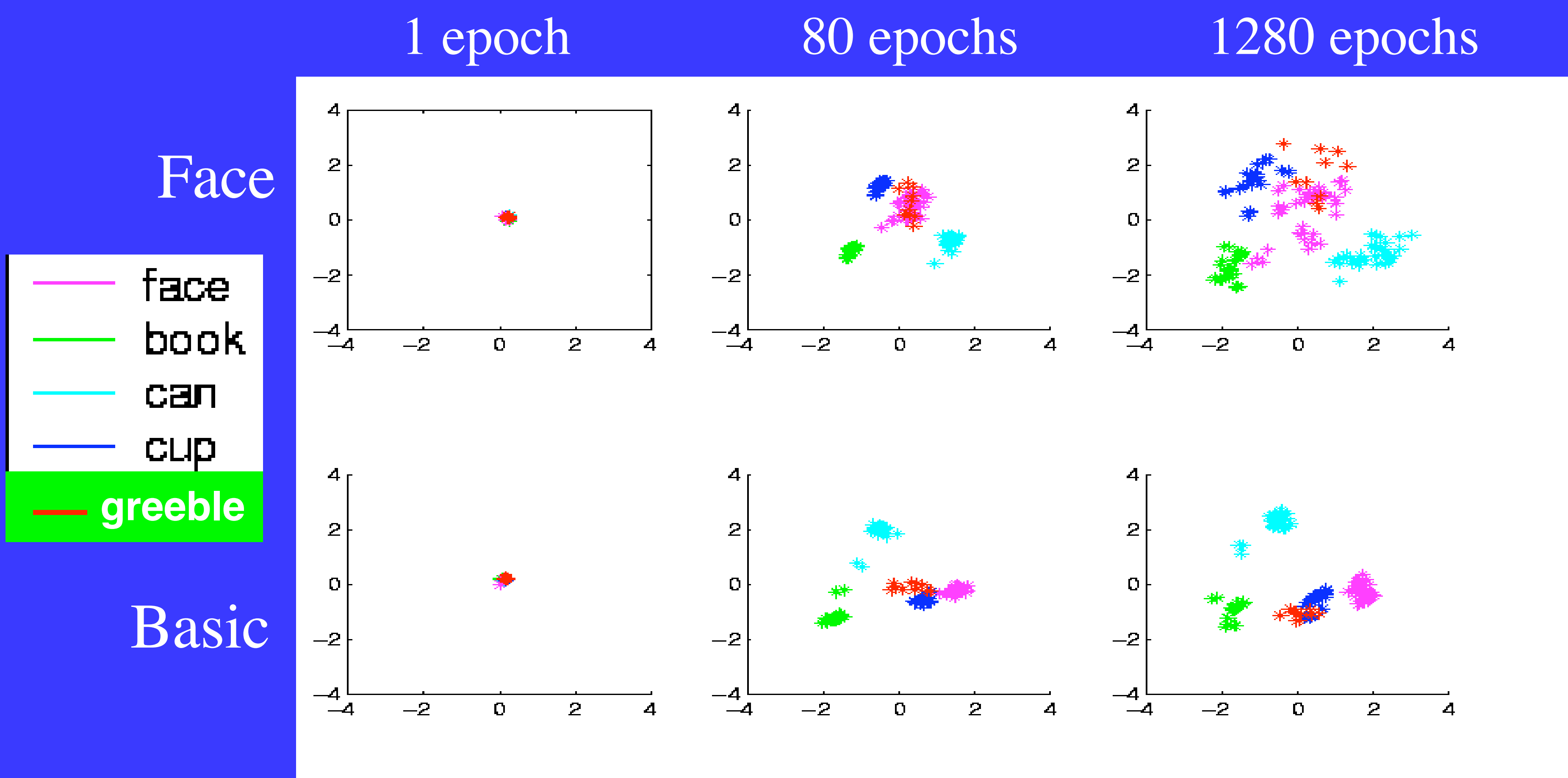
How do experts learn the task?

- Expert level networks must be *sensitive* to within-class variation:
  - Representations must **amplify** small differences
- Basic level networks must *ignore* within-class variation.
  - Representations should **reduce** differences

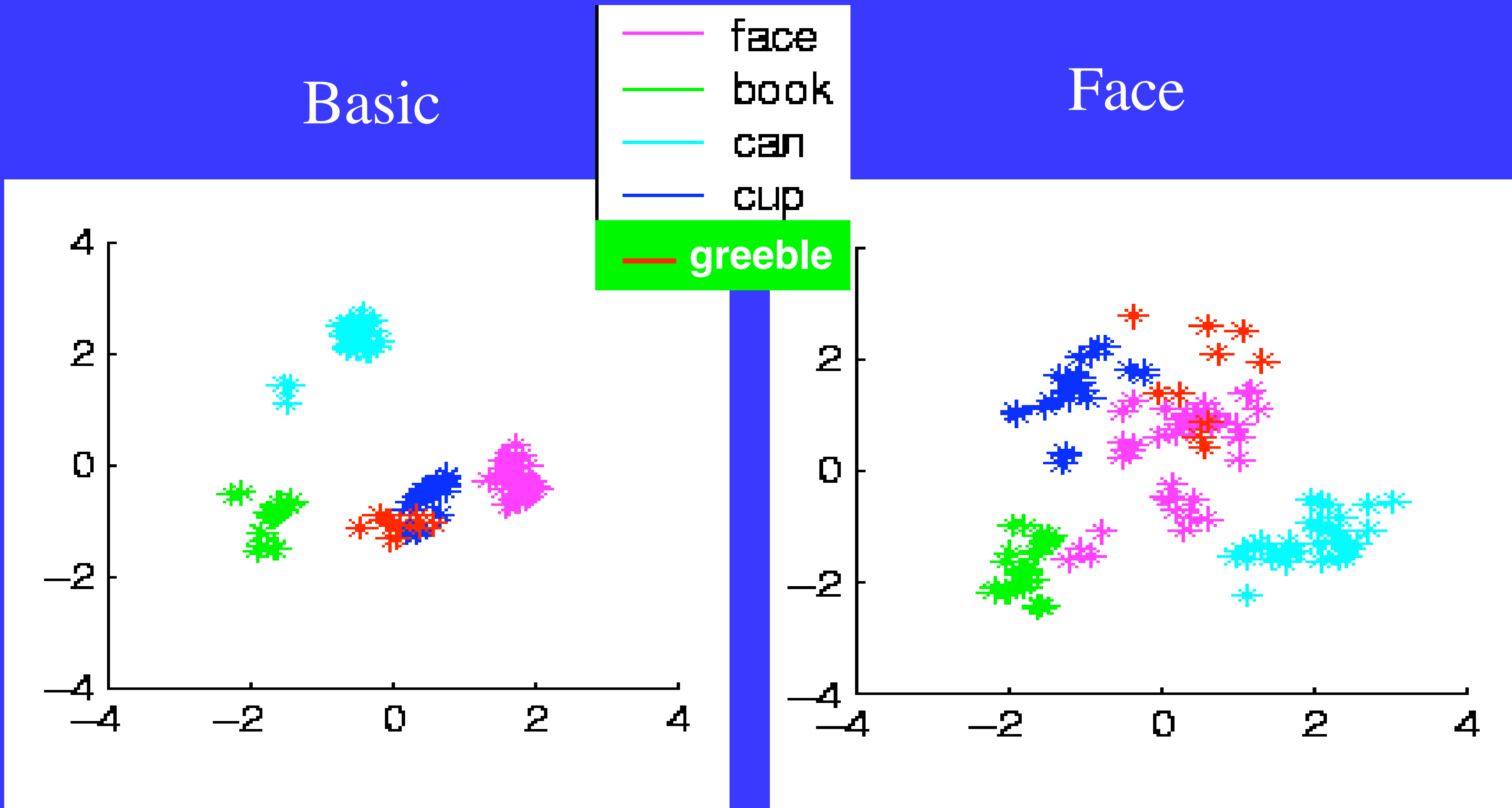
Observing hidden layer representations

- Principal Components Analysis on hidden unit activation:
  - PCA of hidden unit activations allows us to reduce the dimensionality (to 2) and plot representations.
  - We can then observe how tightly clustered stimuli are in a low-dimensional subspace
- We expect basic level networks to separate classes, but not individuals.
- We expect expert networks to separate classes and individuals.

Subordinate level training magnifies small differences *within* object representations

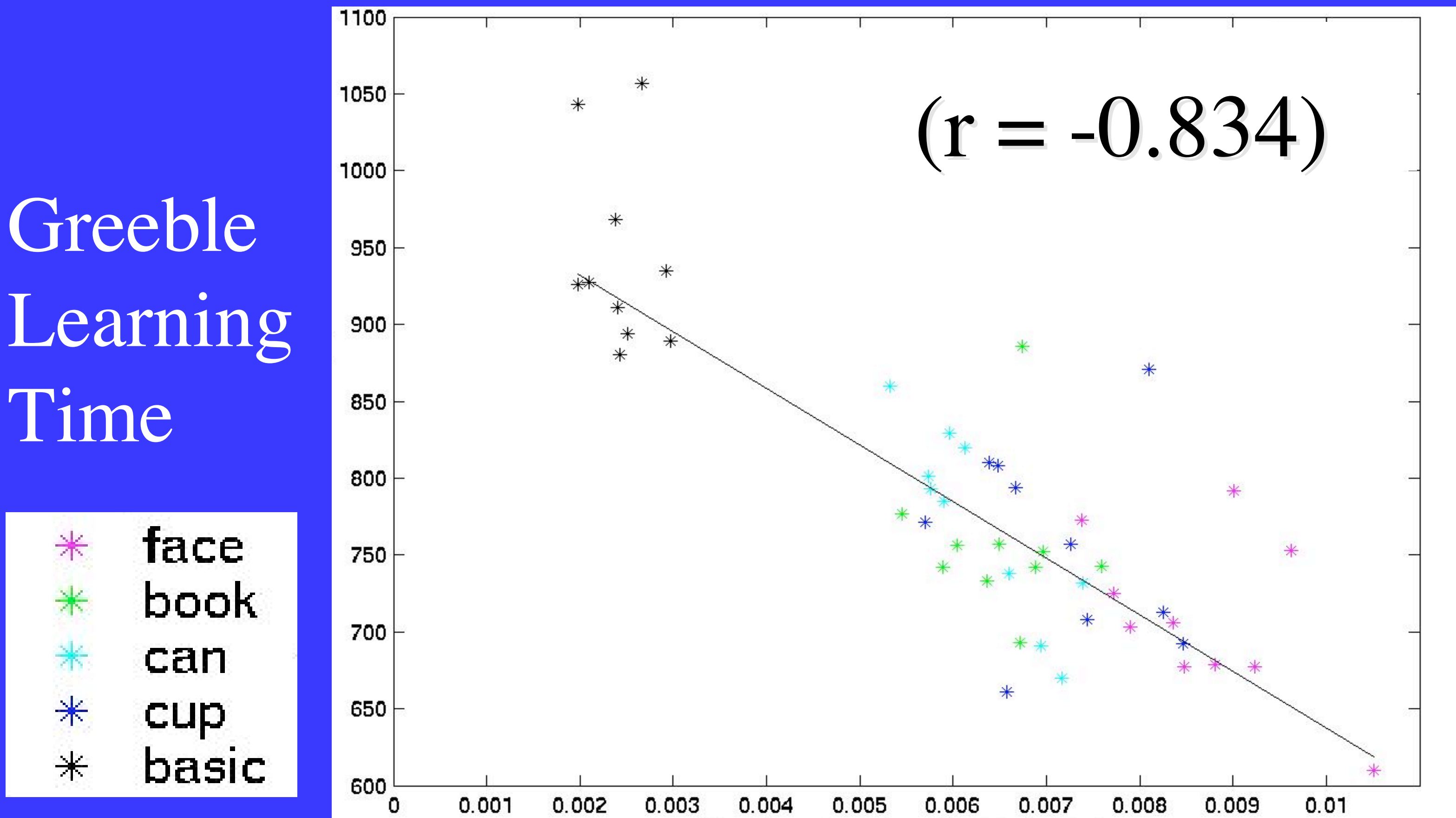


Greeble representations are spread out prior to Greeble Training



Variability Decreases Learning Time

Greeble Learning Time



Greeble Variance Prior to Learning Greebles

Examining the Net's Representations

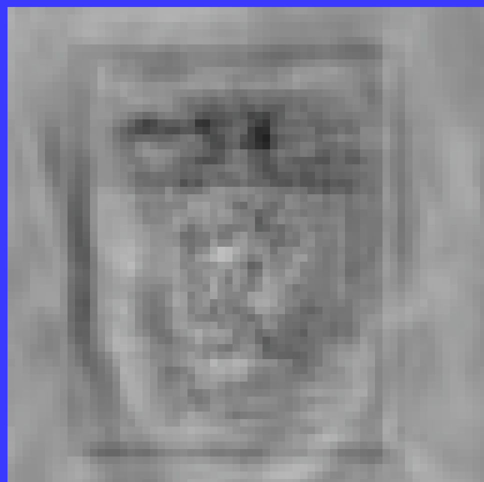
- We want to visualize "receptive fields" in the network.
- But the Gabor magnitude representation is noninvertible.
- We can *learn* an approximate inverse mapping, however.
- We used linear regression to find the best linear combination of Gabor magnitude principal components for each image pixel.
- Then projecting each hidden unit's *weight vector* into image space with the same mapping visualizes its "receptive field."

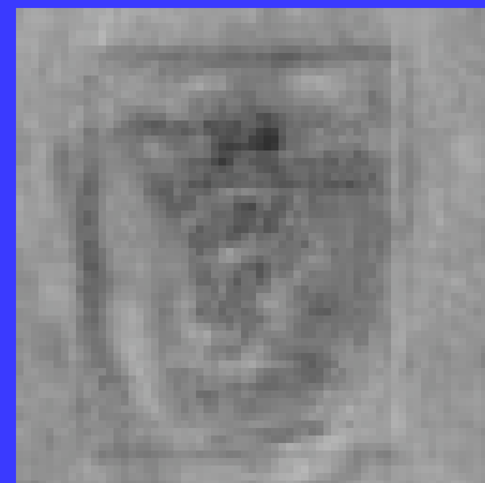
Two hidden unit receptive fields

AFTER TRAINING AS A FACE EXPERT

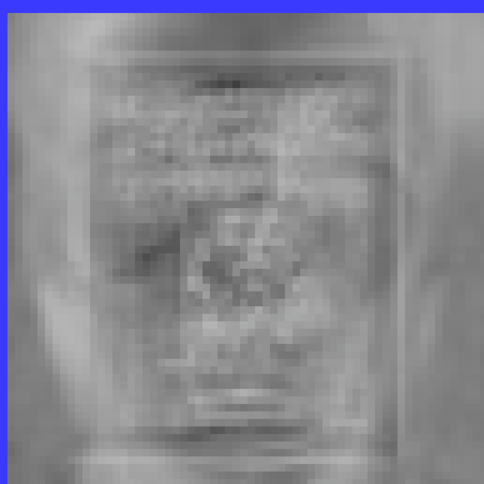
AFTER FURTHER TRAINING AS A GREEBLE EXPER

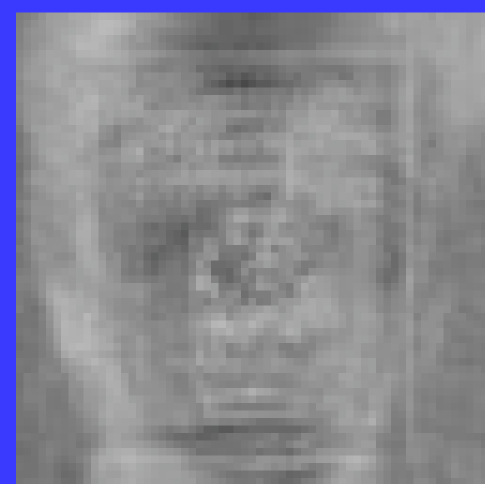
HU 16





HU 36





Conclusion

- Experts learned a new domain of expertise faster.
- The weird thing is: the *more* experts are trained, the *faster* they learn the new task:
  - Suggests the features developed for fine level discrimination (high entropy representations) are good for differentiating other stimuli.
  - Another way to think about it is: for fine level discrimination, similar inputs need to lead to *dissimilar* representations.

Visual expertise is a general skill that is not specific to any class of images including faces.

Wrap up

- We are able to explain a variety of results in face processing.
- Why low spatial frequencies appear to be important in face processing (specialization model: LSF -> better learning and generalization).
- How expression processing can appear to be discrete and continuous at the same time (but it is continuous!).
- Why fear is the hardest expression to recognize.
- Why a face area would be recruited to be a Greeble area: expert level (fine discrimination) processing leads to highly differentiated features useful for *other* discrimination tasks.

END

Conclusions

- The best models perform the same task people do
- Concepts such as “similarity” and “categorization” need to be understood in terms of models that do these tasks
- Our model simultaneously fits data supporting both categorical and continuous theories of emotion
- The fits, we believe, are due to the interaction of the way the categories slice up the space of facial expressions,
- And the way facial expressions inherently resemble one another.
- It also suggests that the continuous theories are correct: “discrete categories” are not required to explain the data.
- We believe our results will easily generalize to other visual tasks, and other modalities.

Outline

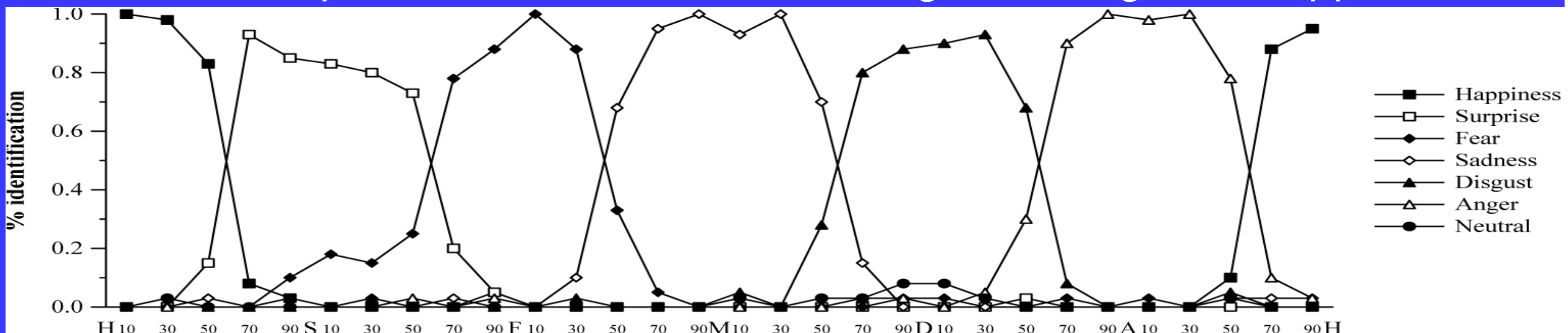
- An overview of our facial expression recognition system.
- The internal representation shows the model's prototypical representations of Fear, Sadness, etc.
- How our model accounts for the "categorical" data
- How our model accounts for the "two-dimensional" data
- Discussion
- Conclusions

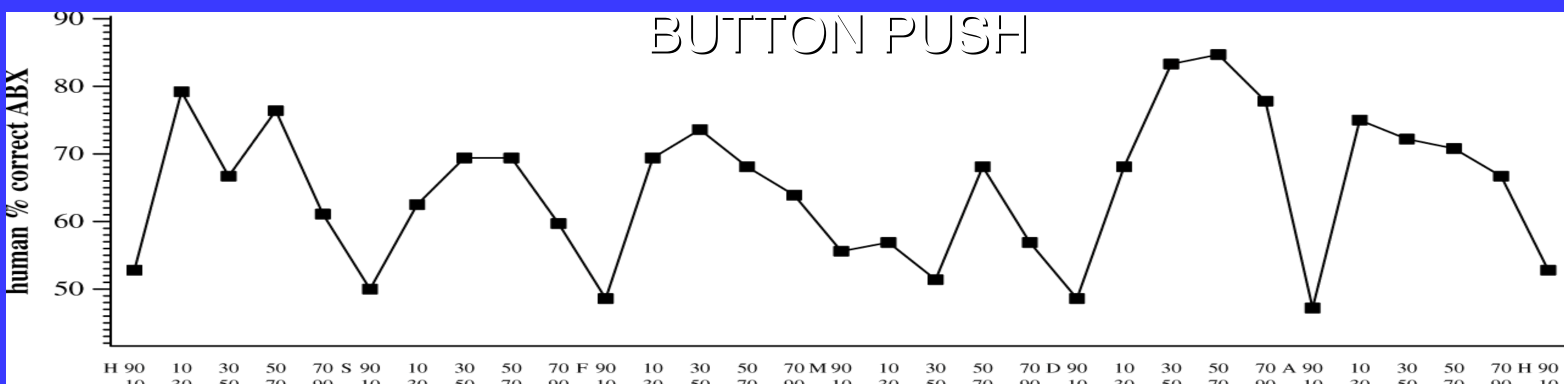
Correlation of Net/Human Errors

- Like all good Cognitive Scientists, we like our models to make the same mistakes people do.
- Networks and humans have a 6x6 confusion matrix for the stimulus set.
- This suggests looking at the off-diagonal terms: The errors
- Correlation of off-diagonal terms: $r = 0.567$. [$F(1,28) = 13.3$; $p = 0.0011$]
- Again, this correlation is an *emergent property* of the model: It was not told which expressions were confusing.

Subject Discrimination Scores

iness Surprise Fear Sadness Disgust Anger Happ-





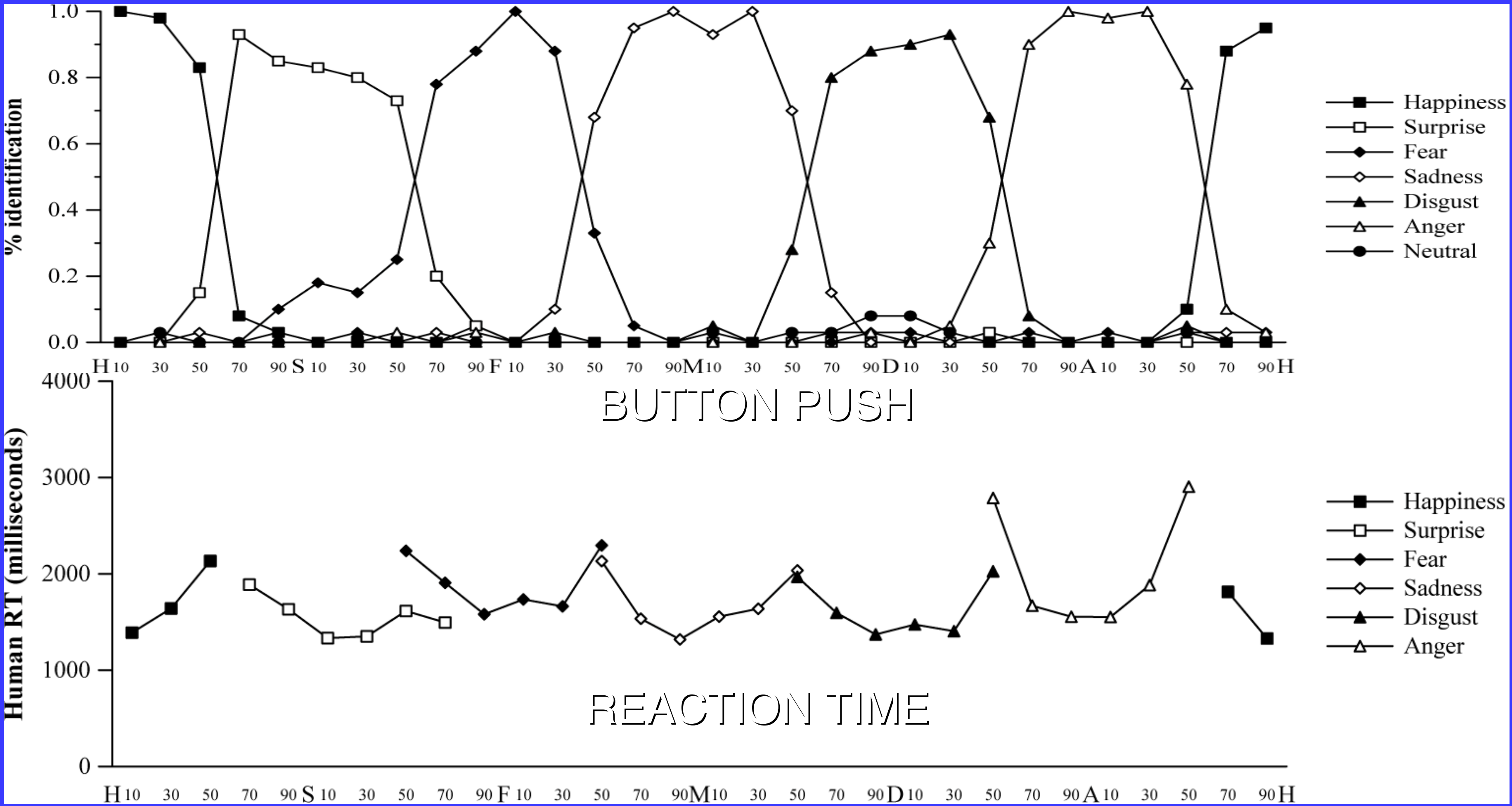
PERCENT CORRECT DISCRIMINATION

- Subjects discriminate pairs of images best when they cross a perceived category boundary

Megamix Human Results

- Sharp transitions, small intrusions, scalloped RTs

iness *Surprise* *Fear* *Sadness* *Disgust* *Anger* *Happ-*



Discrimination

- Classically, one requirement for “categorical perception” is higher discrimination of two stimuli at a fixed distance apart when those two stimuli cross a category boundary
- Indeed, Young et al. found in two kinds of tests that discrimination was highest at category boundaries.
- The result that we fit the data best at a layer before any categorization occurs is significant: In some sense, the category boundaries are “in the data,” or at least, in our representation of the data.

Discussion

- The discrimination correlates with human results most accurately at a precategorization layer: The discrimination improvement at category boundaries is in the representation of data, not based on the categories.
- These results suggest that for expression recognition, the notion of “categorical perception” simply is not necessary to explain the data
- Indeed, most of the data can be explained by the interaction between the similarity of the representations and the categories imposed on the data: Fear faces are similar to surprise faces in our representation – so they are near each other in the circumplex

Discussion

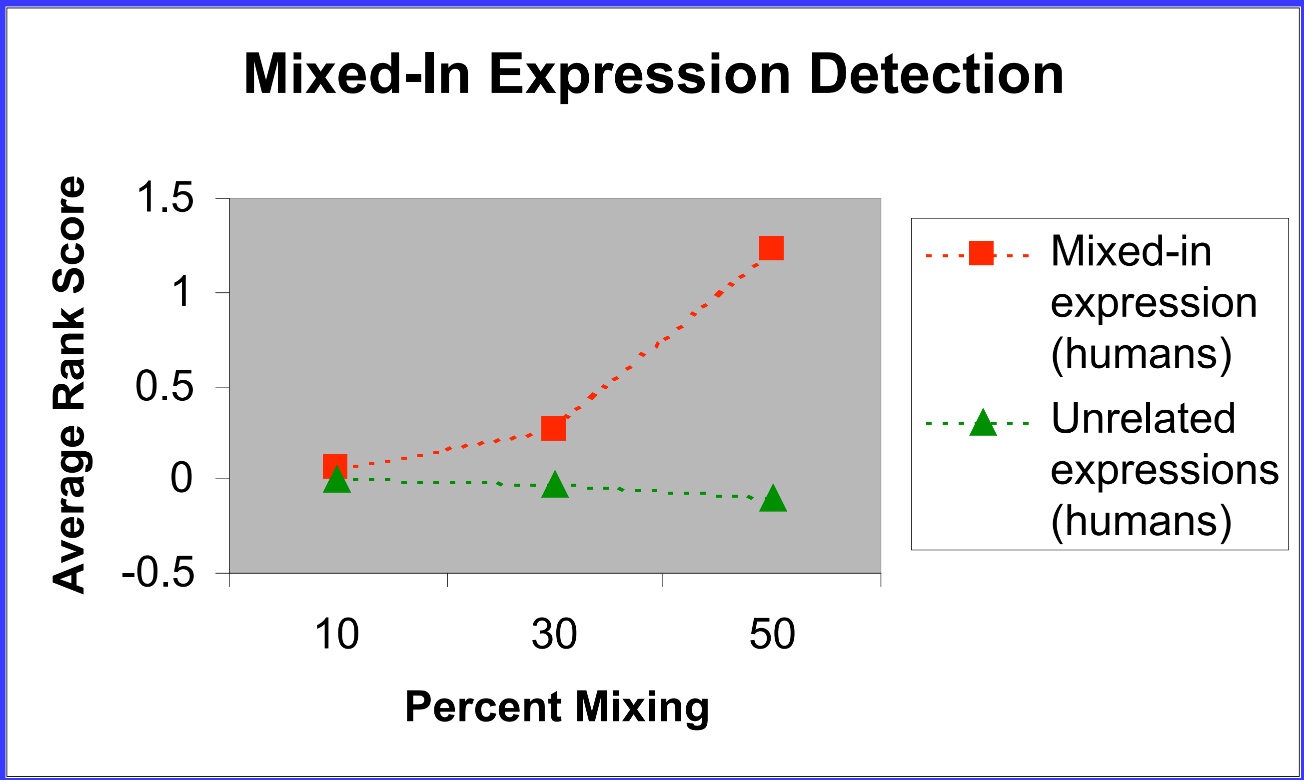
- Our model of facial expression recognition:
  - Performs the same task people do
  - On the same stimuli
  - At about the same accuracy
- Without actually "feeling" anything, without any access to the surrounding culture, it nevertheless:
  - Organizes the faces in the same order around the circumplex
  - Correlates very highly with human responses.
  - Has about the same rank order difficulty in classifying the emotions

Outline

- An overview of our facial expression recognition system.
- How our model accounts for the “categorical” data
- How our model accounts for the “two-dimensional” data
- The internal representation shows the model’s prototypical representations of Fear, Sadness, etc.
- Discussion
- Conclusions

Megamix Human Results

- Young et al. also found evidence for *non*-categorical perception
- Subjects rated 1st, 2nd, and 3rd most apparent emotion.



- At the 70/30 morph level, subjects were above chance at detecting mixed-in emotion. These data seem more consistent with *continuous* theories of emotion.